UNIVERSIDADE FEDERAL DE MINAS GERAIS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS GRADUAÇÃO EM BIOQUÍMICA E IMUNOLOGIA

Tese de Doutorado

Identificação e caracterização de proteínas de ligação a RNA diferencialmente expressas em *Trypanosoma cruzi*

Bruna Mattioly Valente

Belo Horizonte, MG

2018

Bruna Mattioly Valente

Identificação e caracterização de proteínas de ligação a RNA diferencialmente expressas em *Trypanosoma cruzi*

TESE DE DOUTORADO APRESENTADA AO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOQUÍMICA E IMUNOLOGIA DA UNIVERSIDADE FEDERAL DE MINAS GERAIS COMO REQUISITO PARCIAL PARA OBTENÇÃO DO TÍTULO DE DOUTOR EM BIOLOGIA MOLECULAR.

Orientadora: Santuza Maria Ribeiro Teixeira

Belo Horizonte, MG

2018

Agradecimentos

À Deus pela oportunidade de concluir mais um ciclo de aprendizado.

À toda minha família, em especial aos meus pais, ao meu marido e à minha irmã que sempre estiveram ao meu lado me apoiando incondicionalmente em todos os momentos.

À professora Santuza pela oportunidade, pelos ensinamentos, pela dedicação e apoio.

Às alunas de iniciação científica pela ajuda nos experimentos.

A todos meus colegas do laboratório pelos incentivos, pelos momentos de alegria e descontração, pelas conversas e por toda a força que me deram nos momentos mais difíceis. Ao Elimar e a Renata que a todo momento estiveram dispostos a ajudar.

À todos que de alguma maneira contribuíram para esse trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudo.

Agradeço todas as dificuldades que enfrentei; não fosse por elas, eu não teria saído do lugar. As facilidades nos impedem de caminhar. Mesmo as críticas nos auxiliam muito.

Chico Xavier

I. RESUMO

O *Trypanosoma cruzi*, agente causador da doença de Chagas, tem três estágios de desenvolvimento que são bioquimicamente e morfologicamente distintos e que respondem rapidamente às mudanças ambientais que o parasita enfrenta durante seu ciclo de vida. Ao contrário de outros eucariotos, o genoma do *T. cruzi* contêm genes codificadores de proteínas que são transcritos em pré-mRNAs policistrônicos antes de serem processados em mRNAs maduros por meio de reações de *trans-splicing* e de poliadenilação. Sendo assim, o controle da expressão gênica depende, principalmente, de mecanismos pós-transcricionais que podem ser mediados por proteínas de ligação a RNA (RBP), que atuam controlando a estabilidade dos níveis de mRNA e/ou a taxa de tradução desses mRNAs. Pesquisando por motivos presentes em RBPs de eucariotos, identificamos no genoma do clone CL Brener do *T. cruzi*, 253 sequencias codificantes para proteínas contendo motivos de reconhecimento a RNA (RRM), PABP, Alba, Pumilio e o motivo *zinc finger*. Usando dados de RNA-Seq gerados a partir de mRNAs presentes em epimastigotas, tripomastigotas liberadas em cultura e amastigotas coletadas em dois tempos de infecção de fibroblastos humanos, analisamos a expressão das RBPs de *T. cruzi* ao longo do ciclo de vida desse parasita. Entre os cinco genes com expressão aumentada em epimastigotas de CL Brener, comparada com amastigotas e tripomastigotas, identificamos o gene TcCLB.506739.99 que codifica uma RBP contendo motivo *zinc finger*. A importância dessa RBP foi revelada em estudos com epimastigotas nocautes que mostraram redução de crescimento no final da fase logarítimica e um aumento na capacidade de diferenciação em tripomastigotas metacíclicas comparado aos parasitos selvagens. Análises de expressão gênica global (RNA-Seq), a partir de

mRNAs obtidos de epimastigotas nocautes revelaram 12 genes com expressão diminuída nas linhagens nocautes quando comparadas com epimastigotas selvagens. Nenhum gene apresentou aumento de expressão em linhagens nocautes comparado com parasitos selvagens. A capacidade dessa RBP de interagir com o mRNA codificante para a proteína associada a diferenciação, identificado entre os genes diferencialmente expressos no RNA-Seq dos parasitos nocautes, foi confirmada em ensaios de imunoprecipitação de RNA. Sabendo que a população de *T. cruzi* é altamente heterogênea com cepas apresentando diferentes características biológicas, bioquímicas e moleculares, buscamos estudar a expressão de RBPs não somente no clone CL Brener de *T. cruzi*, o qual apresenta alta virulência em modelos de infecção, mas também no clone CL-14, o qual é totalmente avirulento. Para investigar os mecanismos regulatórios que controlam a transcrição de genes que controlam a diferenciação das amastigotas intracelulares em tripomastigotas extracelulares procuramos por RBPs que são diferencialmente expressas durante a infecção de células com os clones CL Brener e CL-14 do *T. cruzi*. Dentre os sete transcritos diferentemente expressos durante o ciclo intracelular, a RBP codificada pelo gene TcCLB.507611.300, contendo o motivo RRM, foi o único que apresentou expressão diferencial entre os clones CL Brener e CL-14. Essa RBP é três vezes mais expressa em tripomastigota de CL-14 comparado à tripomastigota de CL Brener, sugerindo que possa ter um papel regulatório relacionado ao fenótipo não virulento de CL-14.

II. ABSTRACT

Trypanosoma cruzi, the causative agent of Chagas disease, has three biochemically and morphologically distinct developmental stages that are programed to rapidly respond to all environmental changes the parasite faces during its life cycle. Unlike other eukaryotes, the *T. cruzi* genome contains protein-coding genes that are transcribed into polycistronic pre-mRNAs before they are processed into mature mRNAs through coupled trans-splicing and poly-adenylation reactions. Because of this, control of gene expression relies mainly on post-transcriptional mechanisms that must be mediated by RNA binding proteins (RBP) that control steady-state levels and/or translation rates of mRNAs. After searching for motifs present in eukaryotic RBPs, we identified in the *T. cruzi* CL Brener genome, 253 sequences encoding proteins containing RNA recognition motif (RRM), PABP, Alba, Pumillio and zinc finger motifs. Using RNA-seq data generated with mRNA present in epimastigotes, tissue culture trypomastigotes and amastigotes extracted at two time points during infection of human fibroblasts, we analyzed the expression of all *T. cruzi* RBPs throughout the life cycle of this parasite. Among the five genes up-regulated in CL Brener epimastigotes compared with amastigotes and trypomastigotes, we found the gene TcCLB.506739.99 which encodes a RBP containing a zinc finger motif. The importance of the protein encoded for this RBP was revealed by knockdown parasites which showed decrease in the end of logarithmic phase of growth. Null mutants also reveals high capacity of differentiation in metacyclic trypomastigotes compared with wild type epimastigotes. Global gene expression were analyzed by RNA-Seq using mRNA of null mutants and reveals 12 genes with differential expression, but no gene were up-regulated in null mutants compared with wild type parasites. The capacity of this

RBP to bind a mRNA encoding for a protein associated with differentiation, found in RNA-Seq data, were confirmed by immunoprecipitation assays. The *T. cruzi* population is highly heterogeneous with strains presenting different biological, biochemical and molecular characteristics. One example of this is CL Brener and CL-14 cloned *T. cruzi*, that are virulent and avirulent clones respectively. To investigate the regulatory mechanisms controlling the distinct transcriptional programs that drive the development from intracellular amastigotes to the extracellular trypomastigote stage, we searching for RBPs that are differentially expressed throughout the infection of the host cell by CL Brener and CL-14 cloned *T. cruzi*. Among the seven transcripts differently expressed during the intracellular life cycle, the RBP encoded by the TcCLB.507611.300 gene, containing RRM motif, is the only RBP with differential expression between CL Brener and CL-14. This RBP has a 3-fold increased expression of in CL-14 trypomastigotes when compared to CL Brener, suggesting that this RBP may have has a regulatory role related to the for non-virulent phenotype of CL-14.

III. LISTA DE FIGURAS

IV. LISTA DE TABELAS

V. ABREVIATURAS

3’UTR – região 3’ não traduzida, do inglês *3’ untranslated region*

5’UTR – região 5’ não traduzida, do inglês *5’ untranslated region*

CRISPR – do inglês *clustered regularly interspaced short palindromic repeats*

DNA – ácido desoxirribonucléico

DTU – Unidade Discreta de Tipagem, do inglês *Discret Typing Units*

GFP – proteína fluorescente verde, do inglês *green fluorescent protein*

GAPDH – gliceraldeído-3-fosfato desidrogenase

HA - hemaglutinina

HR – recombinação homóloga

Hx1 - 5’UTR do gene TcP2β

mRNA – RNA mensageiro

PABP – proteína de ligação a cauda poli(A), do inglês poli(A) binding protein

PAD – proteína associada a diferenciação, do inglês *protein associated with differentiation*

pI – ponto isoelétrico

pré-mRNA – pré RNA mensageiro

PCR – reação em cadeia da polimerase, do inglês *polymerase chain reaction*

PUF – pumilio

RBP – proteína de ligação a RNA, do inglês *RNA binding protein*

RHS – do inglês *retrotransposon hot spot*

RNA – ácido ribonucléico

RNAi – RNA de interferência

RNA Pol I – RNA polimerase I

RNA Pol II – RNA polimerase II

RRM – motivo de reconhecimento à RNA, do inglês *RNA recognition motif*

rRNA – RNA ribossômico

RT-PCR – PCR em tempo real

sgRNA – do inglês *single guide RNA*

SL – do inglês *spliced leader*

UTR – região não traduzida, do inglês *untranslated region*

SUMÁRIO

1. INTRODUÇÃO

1.1 O *Trypanosoma cruzi* e a doença de Chagas

A doença de Chagas, também conhecida como Tripanossomíase Americana é considerada uma doença tropical negligenciada pela Organização Mundial da Saúde (OMS) que estima haver, ainda no mundo, 8 milhões de pessoas infectadas pelo protozoário *Trypanosoma cruzi* e mais de 25 milhões de pessoas sob o risco de contrair a doença (disponível em: http://www.who.int/chagas/disease/en/). Como pode ser observado na figura 1, a doença de Chagas concentra-se principalmente em países situados na América Latina, considerada área endêmica para a doença, e tem se espalhado, nas últimas décadas, para outros países e continentes. Isso tem acontecido devido, principalmente, à migração de pessoas que habitam áreas endêmicas para outros países do mundo tornando a doença de Chagas uma questão de saúde também relevante para países do continente europeu e da América do Norte (1).

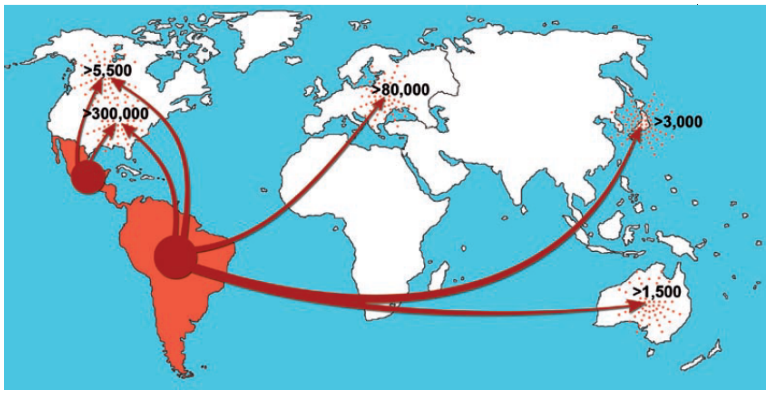


Figura 1: Distribuição geográfica da doença de Chagas. Migração da doença de Chagas de países latino americanos para outros países e continentes. Dados mostram mais de 300.000 pessoas infectadas com o *Trypanosoma cruzi* nos Estados Unidos, 5.500 no Canadá, 80.000 no continente Europeu, 3.000 no Japão e 1.500 na Austrália. Regiões em vermelho indicam áreas endêmicas para doença de Chagas (Coura, 2010).

A transmissão da doença de Chagas ocorre, principalmente, durante o repasto sanguíneo de várias espécies de insetos dos gêneros *Triatoma*, *Panstrongylus e Rhodnius*, encontrados nas regiões endêmicas (2). Sabe-se porém que a transmissão pode ocorrer ainda durante a gestação, por transfusão de sangue, transplante de órgãos, acidentes de laboratório, ingestão de comidas e bebidas contaminadas com o inseto vetor ou com as fezes de triatomíneos contaminados (3, 4). Durante o seu ciclo de vida o *T. cruzi* infecta dois hospedeiros, um vertebrado e um invertebrado, apresentando variadas formas morfologicamente distintas em cada um destes hospedeiros como mostrado na figura 2. Durante o repasto sanguíneo o triatomíneo infectado defeca e urina sob a pele do hospedeiro vertebrado depositando no local da picada as formas infectantes do protozoário, denominadas tripomastigotas metacíclicas, que penetram pelo local da picada ao coçar a pele, ou pelas conjuntivas ao esfregar os olhos. Após invadir diferentes tipos de células nucleadas, as formas tripomastigotas metacíclicas se diferenciam em amastigotas, caracterizadas por serem formas arredondadas com flagelo vestigal, que se multiplicam por divisão binária até se transformarem em tripomastigotas que rompem as células e são liberadas na corrente sanguínea. As tripomastigotas são formas infectantes, que apresentam cinetoplasto posterior ao núcleo e flagelo ao longo do protozoário, que circulam pela corrente sanguínea do hospedeiro vertebrado e infectam outras células dando continuidade ao ciclo celular, ou são ingeridas por triatomíneos durante o hematofagismo reiniciando o ciclo no hospedeiro invertebrado. Quando ingerida pelo triatomíneo as tripomastigotas atingem o estômago do inseto, onde se transformam em epimastigotas que se diferenciam em tripomastigotas metacíclicas ao atingiram a porção final do intestino e são eliminadas pelas fezes (Figura 2) (5, 6).

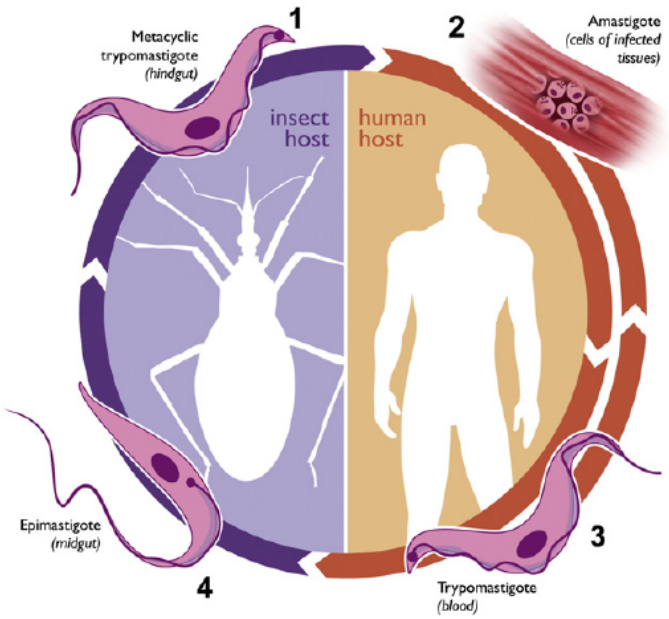


Figura 2: Ciclo de vida do *Trypanosoma cruzi*. 1- Tripomastigotas metacíclicas são excretadas pelas fezes e urina durante o repasto sanguíneo dos triatomíneos e infectam o hospedeiro vertebrado. 2- Após infectar células nucleadas, as tripomastigotas metacíclicas se diferenciam em amastigotas que multiplicam por divisão binária. 3- Amastigotas intracelulares se diferenciam em tripomastigotas que rompem a célula e são liberadas em corrente sanguínea para invadir novas células ou serem ingeridas pelo inseto durante o repasto sanguíneo. 4- Tripomastigotas sanguíneas chegam no estômago do barbeiro onde se transformam em epimastigotas que migram até o intestino onde replicam e se transformam em tripomastigotas metacíclicas dando continuidade ao ciclo de vida do *T. cruzi* (Cuervo et.al, 2010).

A doença de Chagas pode ser dividida em duas fases, aguda e crônica. A fase aguda tem início, aproximadamente, de 2 a 3 semanas após a infecção pelo *T. cruzi* e costuma ser assintomática podendo haver sintomas parecidos aos de uma gripe como: febre, dor muscular e dor de cabeça. Esta fase é marcada pela elevada parasitemia e em casos de transmissão vetorial em que a porta de entrada do parasito no hospedeiro vertebrado ocorre pela conjuntiva geralmente é encontrado o sinal de Romanã, e quando ocorre pela pele é possível observar o chagoma de inoculação. Dois ou três meses após a infecção, estabelece-se a fase crônica da doença que tem como característica baixo número de parasitos na carente sanguínea e baixos níveis de parasitismo tecidual. Em torno de 30% a 40% dos indivíduos infectados apresentam, nesta fase, alterações cardíacas ou gastrointestinais, enquanto que a maior parte dos indivíduos não apresentam sintomas clínicos (7). Os motivos pelos quais alguns hospedeiros desenvolvem a forma sintomática durante esta fase ainda não são conhecidos (8), mas certamente fatores relacionados à variabilidade genética tanto do parasito quanto do hospedeiro devem ser considerados.

Cem anos após a descoberta e caracterização da doença por Carlos Chagas, ainda não há um tratamento eficaz para a doença, uma vez que as drogas atualmente utilizadas, além de muitos efeitos colaterais possuem baixa eficácia na fase crônica (9).

1.2 Estudos de Genômica e variabilidade genética na população do *Trypanosoma cruzi*

A ocorrência de diversas cepas retrata uma estrutura populacional heterogênea deste parasito que apresentam diferenças no tamanho do genoma bem como polimorfismos entre cromossomos de diferentes cepas (10). Inicialmente, dados baseados em características biológicas e bioquímicas assim como técnicas

moleculares permitiram a divisão das diversas cepas do *T. cruzi* em 2 grandes grupos, denominados *T. cruzi* I e *T. cruzi* II (11), que estavam associados com a ocorrência dos ciclos doméstico e silvestre respectivamente (12). Anos depois, estratégias de genotipagem mais acuradas permitiram a classificação das diversas cepas do *T. cruzi* em seis grupos ou DTUs (do inglês, *Discret Typing Units*) denominados *T. cruzi* I-VI (13). O clone CL Brener, que foi escolhido como referência para o projeto genoma do *T. cruzi* pelo fato de ter sido bem caracterizado experimentalmente, pertence ao DTU VI que se caracteriza por ser composto por cepas híbridas oriundas de eventos de hibridização entre parasitos pertencentes aos DTUs II e III. Análises da sequência completa do genoma do *T. cruzi*, publicada em 2005 por El-Sayed e colaboradores, baseada na técnica "*Whole Genome Shotgun*" com cobertura de 14 vezes e montagem final de 838 *scaffolds e* 4.008 *contigs*, revelou que 50% do genoma é formado por sequencias repetitivas (14). Devido à natureza híbrida do clone CL Brener decidiu-se por sequenciar o clone Esmerado, pertencente ao DTU II, a fim de comparar os contigs montados para CL Brener com as *reads* geradas para clone Esmeraldo. Dessa maneira foi possível distinguir os dois haplótipos de CL Brener que foram então denominados *Esmeraldo-like* e *Non-Esmeraldo-like*. O sequenciamento do clone CL Brener revelou ainda a presença de grandes famílias multigênicas, com algumas famílias sendo compostas por mais de 1.000 membros (14). Com o intuito de gerar uma montagem com maior resolução e que representasse melhor os cromossomos do *T. cruzi*, Weatherly e colaboradores (2009) geraram modelos para cada par de cromossomos homólogos baseando-se na sintenia com os 11 cromossomos previamente montados para o *Trypanosoma brucei* (15, 16). Com os *contigs* restantes, foi completada a montagem utilizando o mapeamento de ambas as extremidades de clones de *Bacterial Artificial Chromosome* (BAC) do *T. cruzi*, o que

resultou em uma montagem com 41 cromossomos para o clone CL Brener do *T. cruzi* mas que ainda possui uma grande quantidade de sequencias fragmentadas e que não foram incluídas nessa montagem (15). Mesmo sem ter sido ainda concluída a montagem final do genoma, os dados disponíveis sobre o genoma do clone CL Brener permitiram avanços nos estudos sobre esse parasito, em especial aqueles voltados para a identificação de fatores de virulência, alvos para testes de diagnóstico e vacinas contra a doença de Chagas. Um trabalho recentemente publicado pelo nosso grupo e que foi o assunto da minha Dissertação de Mestrado (veja apêndice I) exemplifica esses avanços (17).

Com o aperfeiçoamento e barateamento das técnicas de sequenciamento e montagem de genomas, o sequenciamento dos genomas de outras cepas do *T. cruzi, como* a de Sylvio X10/1 e a de Dm28c, ambas pertencentes ao DTU I, foram publicados. Além de realizaram o sequenciamento e a montagem parcial do genoma da cepa Sylvio X10/1, Franzén e colaboradores (2011) realizaram uma análise comparativa desse genoma com o do clone CL Brener (18). A montagem revelou que os genomas de Sylvio X10/1 e CL Brener possuem, além de alta sintenia, um conjunto de genes muito similares, apesar de grandes diferenças no que se refere a quantidade de membros pertencentes às famílias multigênicas, e ainda mostraram que a cepa Sylvio X10/1 se assemelha mais ao haplótipo não esmeraldo do clone CL Brener. Já o tamanho do genoma haplóide da cepa Dm28c foi estimado em 27 Mbp, e as análises revelaram que ele se aproxima mais ao genoma da cepa Sylvio X10/1 do que ao do clone CL Brener quando a identidade da sequencia de nucleotídeos é analisada (19).

Um clone do *T. cruzi* derivado da mesma cepa isolada do *Triatoma infestans* (20), da qual foi isolado o clone CL Brener utilizada para iniciar o projeto de sequenciamento, foi denominada CL-14 e apresenta como característica peculiar o

fato de ser um clone avirulento. Essa característica foi observada em ensaios de infecção *in vitro* que mostraram que a taxa de invasão do clone CL-14 em células HeLa é quatro vezes menor quando comparado com a cepa parental CL (21). Através da inoculação em animais de tripomastigotas do clone CL-14, foi demonstrado também que esse clone, incapaz de produzir parasitemia até mesmo em animais imuno-deficientes, induz imunidade protetora ao desafio com a cepa CL, prevenindo a mortalidade, o desenvolvimento de parasitemia e sintomas da doença de Chagas em camundongos (22, 23). Recentemente nosso grupo iniciou análises do genoma do clone CL-14 que indicaram alta similaridade entre os clones CL Brener e CL-14. Por meio da amplificação dos marcadores SL, 24S rDNA e COII em DNAs purificados de culturas de formas epimastigotas de CL-14 e CL Brener, nosso grupo mostrou que assim como CL Brener, CL-14 também é um clone híbrido derivado de cepas pertencentes ao DTU II e ao DTU III, devendo ser classificado com DTU VI (24). Com base em comparações entre as sequências dos genomas dos clones CL Brener, disponíveis no banco de dados TritrypDB (disponível em http://tritrypdb.org/tritrypdb/) e sequências do genoma de CL-14, que foram geradas pelo nosso grupo utilizando a plataforma 454 FLX, verificou-se uma alta similaridade de sequencia e total identidade na organização do genoma mitocondrial (kDNA) e na quantidade de cópias das famílias de multigênicas, apesar das grandes diferenças no comportamente dessa cepa em ensaios de infecção experimental (Mendonça-Neto, dados não publicados).

1.3 Expressão gênica em tripanosomatídeos

Diferentemente dos eucariotos superiores que possuem transcrição monocistrônica de genes codificadores de proteínas catalisada principalmente pela enzima RNA polimerase II (RNA Pol II), que reconhece promotores específicos para cada gene, nos tripanosomatídeos não foram identificadas sequencias com

características de promotores para RNA Pol II. Em tripanosomatídeos, genes codificadores de proteínas organizam-se em unidades de transcrição policistrônica que são transcritos pela RNA Pol II. Apesar da transcrição no tripanosomatídeos ser mediada pela RNA Pol II, dois grupos de genes codificadores de proteínas do *Trypanosoma brucei* são transcritos pela RNA polimerase I (RNA Pol I) (25, 26). Devido à ausência de íntrons nos tripanosomatídeos, os pré-RNAs mensageiros (pré-mRNAs) policistrônicos são formados pelas sequencias codificadoras dos genes, com suas respectivas regiões não traduzidas 5' e 3' (UTR), que separam-se de outros genes através de pequenas sequencias intergênicas (27). Depois de transcritos, os pré-mRNAs policistrônicos são clivados no núcleo, dando origem aos RNAs mensageiros (mRNAs) monocistrônicos maduros. O processamento dos pré-mRNAs policistrônicos ocorre por meio dos processos de *trans-splicing* e poliadenilação. No processo de *trans-splicing* ocorre a adição do *splice leader* (SL), que é transcrito em outro cromossomo e formado por uma sequência de 39 nucleotídeos contendo uma guanina metilada (cap 5'), na extremidade 5' do RNA mensageiro. A poliadenilação é o processo em que vários nucleotídeos de adenina, conhecido como cauda poli-A, são adicionados à extremidade 3' dando origem aos mRNAs maduros que, em tripanosomatídeos, caracterizam-se por conter SL e cauda poli-A (28). Estudos mostraram que esses dois processos não ocorrem de maneira independente, mas que a adição da cauda poli-A no mRNA de um gene é governada pela posição do sítio aceptor da sequência de SL do gene que vem em sequência. Além disso, foi demonstrado que a inserção do SL e da cauda poli-A é guiada por regiões que contêm sequencias ricas em polipirimidinas localizadas nas regiões intergênicas, pois a presença do dinucleotídeo AG, situado posteriormente a essa região rica em motivos citosina e timina, direciona a inserção do SL (29, 30). Após o processamento, os

mRNAs maduros acumulam em diferentes níveis no citoplasma (Figura 3). Uma revisão sobre o processo de expressão gênica e os principais mecanismos de controle foi recentemente publicada pelo nosso grupo em um capítulo do livro *Frontiers in Parasitology* intitulado: “Molecular and Cellular Biology of Pathogenic Trypanosomatids” (veja apêndice II) (31).

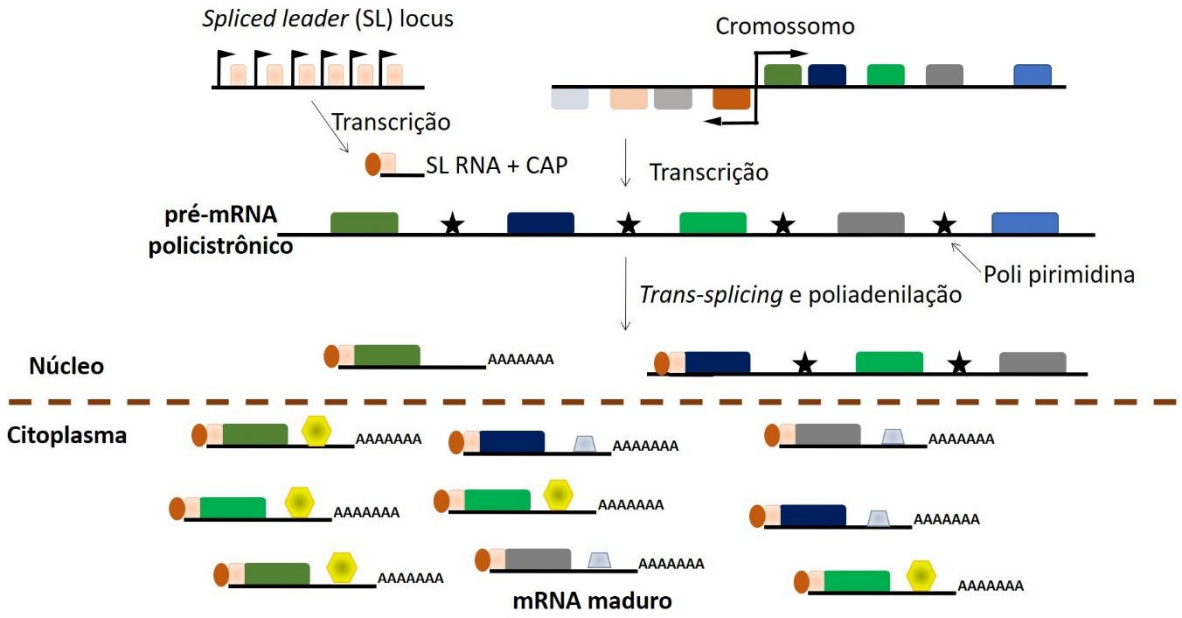


Figura 3: Transcrição e processamento de mRNA em tripanosomatídeos. Vários genes que estão em sequencia no cromossomo formam o pré-mRNA policistrônico e são transcritos ao mesmo tempo. Em outro cromossomo são transcritos genes que codificam o *splice leader* que possui o cap 5' (representados em rosa claro e marrom respectivamente). O trato de poli pirimidina (representado pelas estrelas) guia a adição do *splice leader* no gene abaixo e da cauda poli-A no gene que está acima da sequencia de poli pirimidina através das reações de *trans-splicing* e poliadenilação que ocorrem simultaneamente. Assim os mRNAs maduros são gerados e enviados ao citoplasma onde acumulam-se. Hexágono amarelo e trapézio azul claro representam proteínas de ligação a RNA responsáveis pela regulação pós-transcricional dos mRNAs (Araújo e Teixeira, 2011).

1.4 Proteínas de ligação a RNA (RBP) e o controle da expressão gênica em tripanosomatídeos

Tendo em vista o fato da transcrição do *T. cruzi* ser policistrônica com ausência de sinais conservados envolvidos na regulação da transcrição, onde todos os genes presentes em uma mesma unidade policistrônica são transcritos com a mesma taxa, os estudos sobre a regulação de genes em tripanosomatídeos conduziram a um conceito de regulação pós-transcricional, em que a expressão de um conjunto de genes está coordenadamente regulada ao longo das fases do ciclo de vida, durante o ciclo celular ou em resposta a sinais externos (32, 33). Estudos mostram que as proteínas de ligação a RNA (RBPs) possuem sequencias que se ligam especificamente a motivos regulatórios presentes nos transcritos e têm um papel importante em controlar o acúmulo destes transcritos no citoplasma, seja através da estabilização ou degradação dos mRNAs (32, 33). Interação entre o mRNA e as RBPs ocorrem através do reconhecimento de sequências alvos presentes no mRNA, na maioria das vezes na região 3'UTR, por domínios de ligação presentes nas RBPs, sendo os principais: motivo de reconhecimento de RNA (RRM), proteínas de ligação à cauda poli-A (PABP), motivo dedos de zinco (*zinc finger* – C-X8-C-X5-C-X3-H), domínio Pumilio (PUF) e Alba (34).

O motivo de reconhecimento de RNA (RRM) é altamente conservado entre diversos organismos e composto por aproximadamente 90 aminoácidos com estrutura secundária apresentando topologia de quatro folhas beta antiparalelas e duas alfa hélices (βαββαβ) que apresentam na sua porção central duas sequencias curtas denominadas RNP1 e RNP2 que estão envolvidas com a ligação ao RNA (34, 35). Após a publicação dos genomas de *Leishmania major* (36), *T. cruzi* (14) e *T. brucei* (16), De Gaudenzi e colaboradores (2005) realizaram busca por RBPs contendo

motivos RRM nesses três genomas e identificaram 139 proteínas contendo o motivo RRM em *T. cruzi*, 75 em *T. brucei* e 80 em *L. major* (37). Esses mesmos autores descreveram ainda que apenas 12 RBPs identificas no genoma dos três tripanosomatídeos são conservadas em leveduras e mamíferos, indicando que pode ter ocorrido uma co-evolução entre as RBPs específicas para os tripanosomatídeos e seus alvos presentes no mRNA (37).

Também presentes em RBPs, motivo dedo de zinco ou *zinc finger* é composto por três resíduos de cisteína e um de histidina que coordenam o íon de zinco e ligam-se preferencialmente em regiões ricas em adenina e uracila (38). Pesquisas por proteínas contendo motivos z*inc finger* nos tripanosomatídeos revelou a presença de 131, 233 e 120 RBPs com esse domínio nos genomas de *T. brucei*, *T. cruzi* e *L. major* respectivamente (39). Similarmente aos resultados obtidos em análises comparativas de proteínas contendo motivos RRM (37), Kramer e colaboradores observaram que a maioria das RPBs que possuem motivos *zinc finger* não apresentam similaridade com proteínas de outros eucariotos sendo únicas para os tripanosomatídeos.

O motivo pumilio (PUF) de ligação a RNA é composto por várias repetições de aproximadamente 35 aminoácidos que ficam em sequencia. Em eucariotos a presença de motivos pumilio têm importante papel na regulação da estabilidade e da tradução, interagindo com seus mRNAs alvos geralmente pelo trinucleotídeo UGU (38, 40). Análise do genoma do *T. cruzi* mostrou que no clone CL Brener a família PUF é composta por 10 membros que se dividem em 3 grandes grupos (41).

O motivo de ligação a RNA Alba é composto por várias repetições de arginina e glicina (RGG), sendo que apenas quatro proteínas contendo esse motivo foram identificadas no genoma do *T. brucei* (38, 40).

Ainda que os mecanismos de ação das RBPs sobre os mRNAS, seja promovendo sua estabilização ou degradação e até mesmo dando início a tradução, não estejam completamente esclarecidos, estudos sugerem que as proteínas TbPUF9 (42) e TbZFP3 (43) atuem competindo com fatores de desestabilização sendo que esta última ainda induz a ativação da tradução. De maneira oposta, a proteína de *T. cruzi* TcPUF6 recruta a maquinaria de degradação levando a desestabilização dos mRNAs aos quais ela interage (44). Ao longo do ciclo de vida do parasito, proteínas de ligação a RNA podem sofrer modificações pós-traducionais (como por exemplo serem fosforiladas) em específicos estágios fazendo com que as RBPs se associem com outras proteínas e até mesmo permitindo ou impedindo o acesso da RBP com os transcritos alvos (45). A TcUBP1 descrita como fator de desestabilização e importante na regulação da expressão de mRNA codificante para pequenos genes de mucina (SMUG), é um exemplo de RBP que possui diferentes mecanismos de ação dependo do estágio de vida do *T. cruzi* (46). Em formas epimastigotas, a TcUBP1 interage com as RBPs TcUBP2 e TcPABP1 formando complexos de ribonucleoproteína (mRNP) capazes de estabilizar os transcritos de SMUG, enquanto que nas formas tripomastigotas a TcUBP1 interage apenas com a TcPABP1, devido a ausencia de TcUBP2 neste estágio, desestabilizando mRNAs codificantes para genes SMUG (Figura 4) (47, 48).

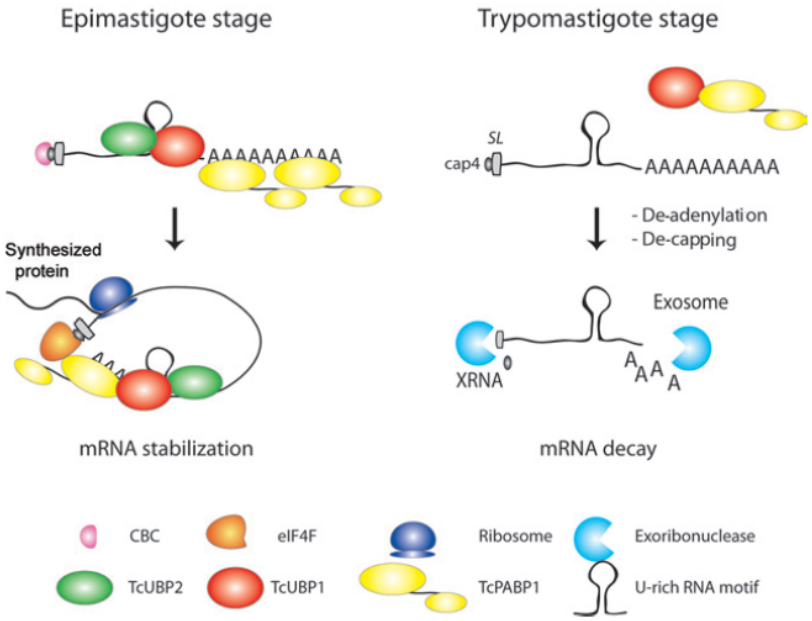


Figura 4: Mecanismo de ação da TcUBP1 em diferentes estágios do ciclo de vida do *T. cruzi*. Em formas epimastigotas a proteína de ligação a RNA, TcUBP1 representada pelo círculo vermelho, interage com as RBPs TcUBP2 (verde) e TcPABP1 (amarelo) formando um complexo de ribonucleoproteína (mRNP) que estabiliza os mRNAs aos quais TcUBP1 interage. Por outro lado, nas formas tripomastigotas a TcUBP1 se liga somente à TcPABP1 levando a desestabilização dos mRNAs (De Gaudenzi *et.al*, 2011).

Uma grande parte dos estudos que buscam estudar o papel de proteínas de ligação à RNA na diferenciação dos diversos estágios do ciclo de vida foram realizados com o *T. brucei*. Diferenciação na mosca tsé-tsé de formas procíclicas não infectivas para a forma infectiva, tripomastigota metacíclica, é dirigida pelo aumento da expressão da TbRBP6 em formas procíclicas (49). As RBPs contendo o motivo *zinc finger*, denominadas TbZFP1 e TbZFP2, foram caracterizadas na transição da forma procíclica para a forma sanguínea. Apesar da TbZFP1 ser expressa apenas na forma procíclica e a TbZFP2 estar presente tanto na mosca tsé-tsé quanto no hospedeiro vertebrado (38), estudos mostram que tanto a redução dos níveis da TbZFP2 quanto a ausência da TbZFP1 inibem a diferenciação das formas sanguínea em procíclica (50, 51). Outra RBP importante no ciclo de vida do *T. brucei* é a TbRBP10 que tem expressão aumentada no estágio sanguíneo regulando mRNAs sabidamente aumentados neste estágio (52).

No *T. cruzi* os estudos de diferenciação ao longo do ciclo de vida têm como principal alvo a diferenciação que acontece no hospedeiro invertebrado, isto é, de formas epimastigota para tripomastigota metacíclica. Entre as RBPs envolvidas na metaciclogênese destaca-se a TcZFP2 que possui o motivo *zinc finger* e expressão diminuída nas formas infectantes tripomastigota metacíclica (53). Mörking e colaboradores (2012) mostraram ainda essa RBP regula, além do seu próprio mRNA, transcritos sabidamente aumentados em tripomastigotas metacíclicas sugerindo que ela atua desestabilizando genes associados com a interação entre o parasito e o hospedeiro. Recentemente, o estudo da TcZC3H31, que é única RBP contendo três dominios *zinc fingers,* mostrou que essa proteína é exclusivamente expressa nas formas presentes no hospedeiro invertebrado, isto é, em epimastigotas e tripomastigotas metacíclicas. Tanto a ausência da expressão, gerada por meio do

nocaute gênico, quanto a super expressão dessa RBP em epimastigotas não afetaram o crescimento dessa forma replicativa encontrada no barbeiro. Por outro lado, experimentos *in vitro* mostraram que parasitas nocautes inibem a capacidade de diferenciação de epimastigotas em tripomastigotas metacíclicas enquanto que os super expressores apresentam aumento da taxa de metaciclogênese. Ensaios *in vitro* revelaram que linhagens nocautes são capazes de colonizar o intestino do triatomínio porém a metaciclogênese foi abolida. Esses autores sugerem que a TcZC3H31 atua como um sensor que responde ao stress nutricional modulando positivamente a estabilidade de mRNAs envolvidos com a metaciclogênese (54). O mRNA correspondente a proteína de ligação a RNA TcPUF6 é expresso nas variadas formas do ciclo de vida do *T. cruzi* (55) sendo que nas formas epimastigotas essa RBP interage e co-localiza com a proteína TcDhh1, compontene da via de degração de mRNA, indicando que a TcPUF6 tem função desestabilizadora (44). Com base nos dados que mostram que os mRNAs associados TcPUF6 apresentam expressão aumentada nas tripomastigotas metacíclicas e que nessa forma infectiva essa RBP não co-localiza com a TcDhh1, Dallagiovanna e colacoradores sugerem que a TcPUF6 tem papel na regulação dos transcritos das tripomastigotas metacíclicas. A TcRBP40 caracteriza-se por interagir com mRNAs codificantes para proteínas contendo domínio transmembrana e assim como a TcZFP2, a sua expressão varia durante a metaciclogênse com níveis reduzidos nas tripomastigotas metacíclicas (56). De maneira interessante a localização da TcRBP40 varia ao longo do ciclo de vida, apresentando-se dispersa no citoplasma das formas amastigotas e tripomastigotas ou concentrada nos reservossomos nas epimastigotas (56). Em *T.cruzi* a TcRP19 foi caracterizada por apresentar expressão aumentada nas formas intracelulares, amastigotas (57). Estudos super expressando essa RBP mostrou que o aumento da

expressão afeta de maneira negativa a infectivade e a metaciclogênese, e ainda que em epimastigotas essa RBP desestabiliza seu próprio mRNA (57, 58). Uma proteína de ligação a RNA contendo o motivo Alba em *T.cruzi*, denominada TcAlba30, apresenta expressão constitutiva ao longo do ciclo de vida do *T. cruzi* e se destaca por interagir com as β-amastias, que são proteínas exclusivas de epimastigotas, regulando de forma negativa os mRNAs codificantes para essa sub família de proteínas de superfície (59).

Recentemente a publicação do transcriptoma de cepa Y do *T. cruzi*, classificado como DTU II, mostrou que dez genes codificadores de RBPs são diferencialmente expressos entre os estágios tripomastigota e amastigota em diferentes tempos de infecção (60). Os dados de Li e colaboradores indicaram que a expressão das RBPs codificadas pelos genes TcCLB.511127.10 e TcCLB.506693.30 estão aumentadas na forma tripomastigota enquanto que a expressão de outras 8 genes contendo motivos de ligação ao RNA apresentam, nas formas amastigotas, acúmulo de transcritos com variados níveis de expressão nos diferentes tempos de infecção analisados.

Uma grande parte dos estudos sobre RBPs, não somente em tripanosomatíeos mas em diversos outros tipos celulares, requerem o uso de metodologias capazes de gerar linhagens nocautes para esses genes, linhagens contendo esses genes com expressão aumentada ou com suas sequencias modificadas. Com essas estratégias de estudos, é possível investigar diretamente o papel dessas proteínas como reguladores da expressão de outros genes, uma vez que ao afetar a sua expressão, a expressão dos seus genes alvos também será afetada.

1.5 Manipulação genética em tripanosomatídeos

Vetores para expressão de genes exógenos em tripanosomatídeos devem conter elementos necessários para o correto processamento dos genes codificantes de proteínas em *T. cruzi* que conforme descrito anteriormente é realizado de maneira peculiar. Desta maneira o vetor precisa ter sinais para adição do SL, que deve estar situado na região antes do códon de início da sequencia codificadora, possibilitando assim a ocorrência do *trans-splicing* e sinais para a poliadenilação, que a ocorre após o término da região codificadora. Usando o gene repórter de luciferase, nosso grupo avaliou o papel de diferentes regiões 5' não traduzidas (5' UTR) e de regiões intergênicas contento sítio de adição de SL e verificou que a sequencia correspondente a 5'UTR do gene TcP2β (Hx1) juntamente com a região intergênica e com a região 3' não traduzida (3' UTR) do gene GAPDH aumentaram de forma significante a expressão do gene repórter em formas epimastigotas (61). Tendo em vista que sequencias promotoras capazes de reconhecer a RNA Pol II não foram identificadas no sequenciamento do genoma do clone CL Brener, sequencia correspondente ao promotor de RNA ribossômico (rRNA), que é reconhecido pela RNA Pol I, são usadas em vetores para expressão de um gene exógeno em *T. cruzi* uma vez que são capazes de produzir um grande aumento na taxa de eficiência da transcrição (26). Com os dados acima, DaRocha e colaboradores (2004) desenvolveram o vetor de expressão pROCK (Figura 5A) que contêm, além de todos os elementos descritos acima, sequencia correspondente ao fragmento da região codificadora do gene de β-tubulina. A presença da sequencia de β-tubulina permite a integração do vetor no *locus* de α-β tubulina do genoma do parasito através da recombinação homóloga (HR) (Figura 5B).

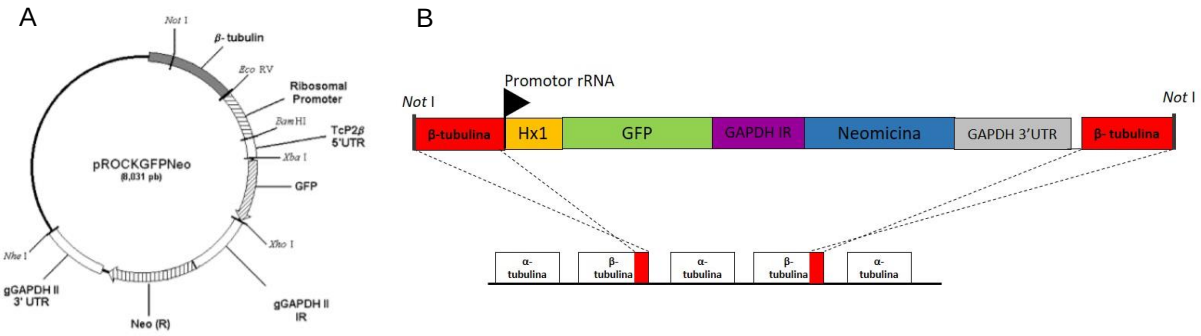


Figura 5: Representação esquemática do plasmídeo pROCK. (A) O plasmídeo de expressão em *T. cruzi* pROCK é composto por sequencia correspondente ao gene de β-tubulina, promotor de rRNA, região 5'UTR do gene Hx1, região codificadora do gene de fluorescência verde (GFP), região intergênica e 3'UTR de GAPDH flanqueando a região codificadora correspondente ao gene de resistência a neomicina. Sequencia correspondente a GFP pode ser removida com as enzimas *Xba*I e *Xho*I para adição da sequencia de outro gene (DaRocha *et al*; 2004). (B) Representação da integração do plasmídeo pROCK_Neo no *locus* de α-β tubulina através do evento de recombinação homóloga.

Estudo da função de um gene pode ser realizado através da redução na expressão do mRNA, pelo silenciamento do gene por interferência de RNA (RNAi), ou pela eliminação completa da expressão do gene que pode ser realizada interrompendo a região codificadora do gene alvo. Transfecção de formas epimastigotas do *T. cruzi* com RNA de dupla fita (dsRNA) de amastina mostrou que não houve a produção dos pequenos RNAs de interferência (siRNA) e nem alteração nos níveis de mRNAs codificantes para as amastinas, indicando que *T. cruzi* não possui todos os componentes necessários para o funcionamento da maquinaria de RNAi (62). Esse dado foi posteriormente confirmado com o sequenciamento do genoma (14) no qual foi constatado que o clone CL Brener não possui sequencias codificadores dos componentes da maquinaria de RNAi. Desta maneira estudos que tem o intuito de investigar a função de um gene em *T. cruzi* utilizam a técnica de interrupção gênica por recombinação homóloga, na qual são utilizados plasmídeos contendo gene de resistência a droga flanqueado por partes homólogas à sequencia do gene alvo (63, 64). Considerando que o *T. cruzi* é um organismo diplóide, nocaute de genes de cópia única por recombinação homóloga exige duas rodadas de transfecção usando duas diferentes drogas para seleção (Figura 6) (65). Apesar de demandar tempo para seleção da população e de clones, essa abordagem relativamente simples para deleção de genes de cópia única torna-se inviável para deleção de famílias multigênicas, uma vez que há limitação no número de marcadores de resistência (66).

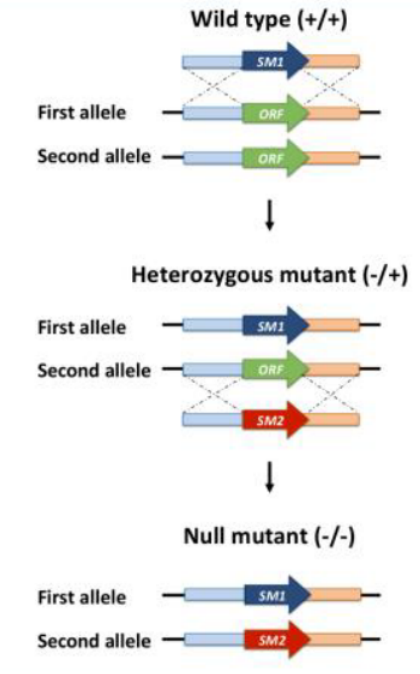


Figura 6: Representação esquemática do nocaute gênico por recombinação homóloga. Parasitos selvagens são transfectados com sequencia homóloga ao gene alvo flanqueando o gene de resistência, representado pela seta azul SM1, resultando na troca da região codificadora do primeiro alelo pela sequencia com o gene de resistência. Após a seleção, parasitos com um dos alelos interrompidos é submetido à transfecção com sequencia contendo outro gene de resistência (SM2, representado pela seta vermelha) flanqueando regiões homologas ao gene alvo para que ocorra a recombinação homologa com o segundo alelo. Parasitos nocautes são obtidos selecionando parasitos resistentes aos dois marcadores (Burle-Caldas *et.al*, 2015).

Recentemente, a tecnologia baseada em repetições palindrômicas curtas (CRISPR) e endonucleases de DNA da proteína Cas associada, denominado CRISPR/Cas, tem sido amplamente utilizada para edição de genomas uma vez que aumenta a eficiência da transfecção ao induzir quebra sítio específica na dupla fita do DNA. O sistema CRISPR/Cas consiste em um complexo ribonucléico, formado por um guia de RNA complementar (sgRNA) ao alvo e a endonuclease Cas, que é capaz de se dirigir até a sequencia alvo e gerar quebra na dupla fita do DNA que pode ser reparada por micro homologia ou favorecer o reparo por recombinação homóloga quando uma sequencia doadora é fornecida (Figura 7) (67). Recentemente Peng e colaboradores (2014) demonstraram que além da edição de genes de cópia única, essa técnica permite a edição de famílias multigênicas (68). Edição gênica utilizando o sistema CRISPR/Cas já foi demonstrado em *T. cruzi* (68, 69) e em outros tripanosomatídeos como *Leishmania donovani* (70) e *Toxoplasma gondii* (71).

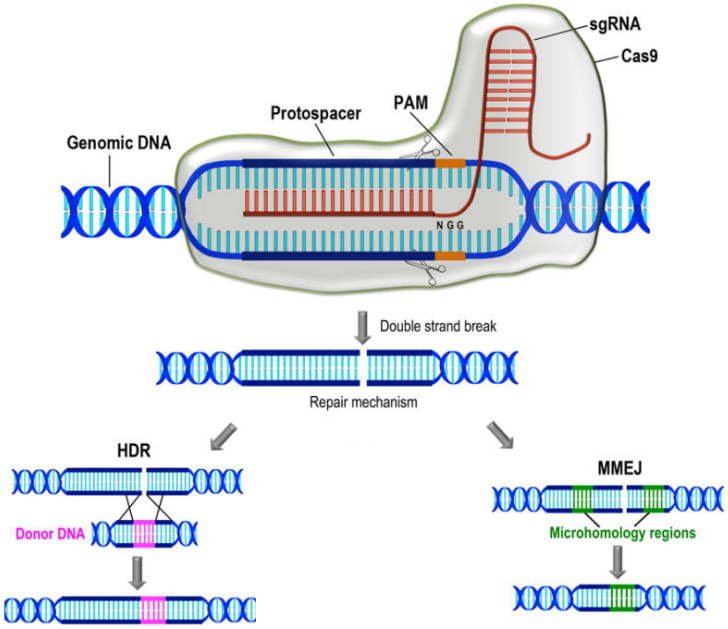


Figura 7: Representação esquemática do sistema CRISPR/Cas. Ligação do sgRNA com a Cas forma o complexo de ribonucleoproteína. O sgRNA reconhece especificamente a região do DNA, que deve estar localizado após a sequencia PAM (varia para cada bactéria), e por complementariedade se liga a ela (representado pelo protospacer). A Cas9 induz quebra na dupla fita do DNA, como indicado pela tesoura, que pode ser reparado por recombinação homóloga (HDR) ou micro homologia (MMEJ), no caso de *T. cruzi*. Reparo da dupla fita do DNA por HDR acontece na presença de uma sequencia que possui homologia à região clivada, enquanto que o reparo por MMEJ ocorre por meio de microrregiões de homologia presentes nas extremidades da fita clivada que se alinham para guiar o reparo, causando pequenas deleções na região da quebra (Adaptado de Lander *et.al*, 2016).

Devido à incapacidade do funcionamento da maquinaria de RNAi e a dificuldade de obter parasitos nocaute para o *T. cruzi*, foi publicado, recentemente, apenas um trabalho, em que a função de uma proteína de ligação a RNA de *T. cruzi* foi investigada por nocaute gênico.

2. OBJETIVOS

2.1 Objetivo geral

Identificar e caracterizar genes codificantes para proteínas de ligação a RNA diferencialmente expressos ao longo do ciclo de vida de clones distintos do *Trypanosoma cruzi*, buscando entender como elas atuam na regulação da expressão de genes estágio-específicos nesse parasito.

2.2 Objetivos específicos

- Identificar sequencias que possuem domínios de ligação a RNA no genoma do *T. cruzi*;
- Determinar os perfis de expressão de genes codificadores para proteínas de ligação a RNA (RPB) ao longo do ciclo de vida dos clones CL Brener e CL-14 do *T. cruzi*;
- Determinar a localização celular das RBPs com expressão aumentada em formas epimastigota de CL Brener e de RBPs diferencialmente expressas entre os clones CL Brener e CL-14;
- Gerar linhagens nocautes de *T. cruzi* para os genes codificantes de proteínas de ligação a RNA e linhagens com níveis de expressão dos mRNAs codificantes para as RBPs aumentados;
- Determinar alterações na expressão gênica global em epimastigotas nocautes para uma RBP com expressão aumentada nesse estágio do ciclo de vida do clone CL Brener;
- Investigar a formação de complexos proteicos envolvendo as proteínas de ligação ao RNA com expressão diferencial no *T. cruzi* por meio de ensaios de co-imunoprecipitação proteína-RNA;

- Identificar motivos na região 3’ UTR, responsáveis pela interação da RBP com o mRNA.

3. METODOLOGIA

3.1 Obtenção e digestão de DNA plasmidial

DNA plasmidial foi extraído utilizando o kit illustra plasmid Prep Mini Spin (GE Healthcare) seguindo as instruções do fabricante. DNA foi digerido incubando-o com enzimas de restrições de acordo com a temperatura e o tempo indicados pelo fabricante.

3.2 Purificação e clonagem

Após a confirmação da clivagem do DNA, por eletroforese em gel de agarose a 100V por aproximadamente 40 minutos, a banda correspondente ao inserto de interesse foi cortada e purificada utilizando o kit NucleoSpin Gel and PCR Clean up (Macherey-Nagel - MN) seguindo as instruções do fabricante. Insertos de interesse foram clonados no vetor de clonagem TOPO 2.1 (TOPO TA Cloning Kit - Invitrogen), incubando o inserto com o TOPO 2.1 durante 30 minutos em temperatura ambiente de acordo com as instruções do fabricante. Clonagens no vetor de expressão em *T. cruzi*, pROCK (61) foram realizadas incubando o inserto de interesse com o plasmídeo pROCK, previamente digerido com *Xba*I e *Xho*I, juntamente com 1 unidade da enzima T4 DNA ligase (T4 DNA ligase- Promega) a 16ºC por 18 horas.

3.3 Transformação de bactérias

Produtos de clonagem foram incubados com bactérias quimiocompetentes E*scherichia coli* (cepa XL1 Blue) por 30 minutos no gelo. Após a incubação foi realizado choque térmico a 42ºC por 55 segundos e em seguida 400µL do meio 2XYT (1,6% de triptona; 1% de extrato de levedura; 0,5% de NaCl; pH 7,0) foi adicionado às bactérias transformadas, que foram então mantidas sob agitação a 37ºC. Após uma hora, aproximadamente 200µL da cultura foram plaqueadas em meio 2XYT-ágar

(meio 2XYT acrescido de 1,5% de ágar) contendo 100µg/mL de ampicilina. As placas foram incubadas por, pelo menos,16 horas a 37ºC.

3.4 PCR de DNA plasmidial/genômico e PCR de colônia

Colônias transformadas foram analisadas por PCR de colônia. Para isso colônias foram incubadas em 20µL de água por 5 minutos e em seguida aproximadamente 2µL foram usados em reação de PCR contendo: 0,2mM de dNTPs, tampão Taq 1X , 1µM de cada iniciador, Taq 0,04U/µL. A reação foi submetida ao programa: 95ºC por 5 minutos seguido de 30 ciclos de 95ºC por 30 segundos, 50ºC por 1 minuto, 72ºC por 2 minutos. Após todos os ciclos as reações eram incubadas a 72ºC por 10 minutos para complementação da extensão das cadeias. Por fim produtos de PCR foram analisados por eletroforese em agarose 1% a 100V por aproximadamente 40 minutos.

PCR utilizando DNA plasmidial ou genômico foram realizadas com 10ng de plasmídeo ou 100ng de DNA genômico, respectivamente. DNAs foram utilizados na mesma reação de PCR e ao mesmo programa descrita acima, variando a temperatura de anelamento e o tempo de extensão, de acordo com a combinação dos iniciadores utilizados. Na tabela 1 encontram-se os iniciadores utilizados neste trabalho.

Tabela 1 – Descrição dos iniciadores usados no trabalho

Iniciador	Sequencia
For_TcRBP99_XbaI	5’tctagaATGCCCAAGGCCAACAAG’3
Rev_TcRBP99::HA_XhoI	5’ctcgagTTAAGCGTAATCTGGTACGTCGTATGGGTATTCGTAATATGTAGAGAGAG’3
For5’KO_TcRBP99_HindIII	5’aagcttGGAGAACTTAAGGCATTCAT’3
Rev5’KO_TcRBP99_SacI	5’gagctcCGTTCAGAGAACTCTCTG’3
For3’KO_TcRBP99_XhoI	5’ctcgagGAGTTGTTGGGTTTGAGC’3
Rev3’KO_TcRBP99_XbaI	5'tctagaCCTGTAGGAAAGCAAAGCCG’3
For_TcRBP99	5’AACTCGCGTTACAAGGAGGG’3
Rev_TcRBP99	5’TCTCCCCCGTAAACCGACCT’3
For_TcRBP300_NheI	5’gctagcATGTTTCTAGAGGAGCTT’3
Rev_TcRBP300::HA_XhoI	5’ctcgagCTAAGCGTAATCTGGTACGTCGTATGGGTAAGCCTGTTTCAAATATAA’3
For5’KO_TcRBP300_KpnI	5’ggtaccTTCTTGTGAGAAATGTGTCT’3
Rev5’KO_TcRBP300_SacI	5’gagctcCTGAGCAGATGCAGATTCAG’3
For3’KO_TcRBP300_XhoI	5’ctcgagTGCCGAGTGACGTATAGACG’3
Rev3’KO_TcRBP300_XbaI	5’tctagaACAGAGAATTCTTAGTTCAG’3
For_sgRNA_TcRBP300	5’GGAGGCCGGAGAATTGTAATACGACTCACTATAGGAGTATCACTACAGCTCCAGCTTCGGATGTTTTAGTACTCTGGAAACAGAATC3’
Rev_sgRNA	5’-AAAAAAATCTCGCCAACAAGTTGAC-3’
For_RpL9_RT.PCR	5’TGACAACTCGACCATCAACA’3
Rev_RpL9_RT.PCR	5’GGCGAAGCGAATCTTAAAAC’3
For_TcRBP99_RT.PCR	5’GGCGAAGCGAATCTTAAAAC’3
Rev_TcRBP99_RT.PCR	5’GGCCTCACGCTCATCTAGTC’3
For_PAD_RT.PCR	5’GGTGGAGACGGACGTCGA’3
Rev_PAD_RT.PCR	5’TAGACAAAGCTGGCGTTG’3
For_transportador_RT.PCR	5’CTGCTGGGCAGCTTCTGT’3
Rev_transportador_RT.PCR	5’GACAACCGCAACCACACCAG’3
For_RHS_RT.PCR	5’ATGGGAATGGAGGTGAGG’3
Rev_RHS_RT.PCR	5’GTACGGCCATCCCTTGTC’3
For_Neo	5’CGACCCTGCAGCCAATATGGGATCG’3
Rev_Neo	5’TCAGAAGAACTCGTCAAGAAGGCG’3
For_Higro	5’ATGAAAAAGCCTGAACTCACCGC’3

For_CDS_PAD	5’ATGACCGGCGAACAATTTG’3
Rev_CDS_PAD	5’TTACGGACGTGCCACCTT’3

*As letras minúsculas indicam sequências referentes à sítios de enzima de restrição.
*As letras sublinhadas representam a sequencia de hemaglutinina (HA).

3.5 Identificação de Proteínas de ligação a RNA (RBP) do *T. cruzi*

Buscas por motivos de ligação ao RNA foram realizadas no banco de dados Pfam (disponível em: http://pfam.xfam.org/), no qual foram identificados 13 domínios (Tabela 2). Todas as sequencias contendo esses domínios depositadas neste banco de dados, foram usadas em para identificar as sequencias codificadoras de proteínas presentes no genoma do clone CL Brener do *T. cruzi*, versão 9.0 depositada no banco de dados TritrypDB (disponível em: http://tritrypdb.org/tritrypdb/), que possuem esses domínios de ligação ao RNA. Para isso análises de homologia entre os motivos de ligação a RNA e as proteínas anotadas de CL Brener foram realizadas através da opção blasp do programa Blast (Basic Local Alignment Search Tool), sendo considerada proteína de ligação a RNA as sequencias com *e-value* menor que 10^{-9} e identidade igual ou superior a 85%.

Tabela 2 – Domínios de ligação a RNA com respectivos identificadores Pfam

Domínio de ligação a RNA	Identificador Pfam
PAP	PF00658
RRM	PF10378
RRM_1	PF00076
RRM_2	PF04059
RRM_3	PF08777
RRM_4	PF10598
RRM_5	PF13893
RRM_6	PF14259
RRM_7	PF16367
PUF	PF00806
Alba	PF01918
Zinc Finger	PF00642
RNA binding domain	PF08675

3.6 Identificação das RBPs diferencialmente expressas

Dados de RNA-Seq das formas epimastigotas, cultivada em meio LIT, amastigotas intracelulares, obtidas 60 e 96 horas após a infecção de *human foreskin fibroblasts* (HFF), e de formas tripomastigotas liberadas em cultura dos clones CL Brener e CL-14 foram gerados em colaboração com o professor Najib El-Sayed (24). Transcritos codificantes para proteínas de ligação a RNA foram comparados nos estágios epimastigota, amastigota 60 horas e 96 horas após a infecção e tripomastigota e os que apresentavam valor de log2 de *fold change* maior que 2 e *p-value* ajustado menor que 0,05, em pelos menos uma das comparações analisadas, foram considerados proteínas de ligação a RNA diferencialmente expressas durante o ciclo de vida do *T. cruzi*.

3.7 Alinhamentos de sequencias e análises filogenéticas

Sequencias de proteínas ortólogas à RBP de interesse de diversos tripanosomatídeos foram extraídas do banco de dados TriTrypDB e submetidas ao programa de alinhamento de múltiplas sequencias Muscle (http://www.ebi.ac.uk/Tools/msa/muscle/). Árvore filogenética foi construída utilizando o programa MEGA (versão 6.0).

3.8 Cultivo de parasitos

Culturas de parasitos dos clones CL Brener e CL-14 do *Trypanosoma cruzi* foram utilizados neste projeto. As formas epimastigotas foram mantidas em meio LIT, *liver infusion tryptose*, (10g/L de infuso de fígado; 4g/L NaCl; 400 mg/L KCl; 8g/L Na2HPO4; 1g/L glicose; 5g/L triptose; 10mL de hemina a 0,2% em NaOH (p/v); pH 7,2), suplementado com 10% de soro fetal bovino inativado, penicilina e estreptomicina a 100U/mL a 28ºC (72). Parasitos transfectados foram mantidos em cultura com 200µg/mL de antibiótico de seleção.

3.9 Construção dos vetores

3.9.1 Vetor para super expressar a TcRBP99

A super expressão do gene TcCLB.506739.99 (TcRBP99) foi realizada utilizando o plasmídeo pROCK (61). A região codificadora dessa RBP foi amplificada do DNA genômico do clone CL Brener utilizando os iniciadores For_TcRBP99_Xba e Rev_TcRBP99:HA_Xho, descritos na tabela 1. Sequencia referente a *tag* de hemaglutinina (HA) foi adicionada à sequencia do iniciador de maneira que esta *tag* fosse adicionada à porção C-terminal durante a amplificação. O produto amplificado foi clonado no vetor TOPO2.1 que em seguida foi digerido com as enzimas de restrição *Xba*I e *Xho*I. Simultaneamente, o plasmídeo pROCK_GFP_Neo foi digerido com estas mesmas enzimas para a remoção da sequência correspondente a proteína verde de fluorescência (GFP). Em seguida a região codificadora da TcRBP99 contendo *tag* de HA foi ligada a este vetor gerando o plasmídeo pROCK_TcRBP99::HA_Neo. Após ser linearizado com a enzima de restrição *Not*I, este plasmídeo foi usado para transfectar formas epimastigotas do clone CL Brener.

3.9.2 Vetor para nocaute do gene que codifica a TcRBP99

O gene TcCLB.506739.99 foi interrompido utilizando o plasmídeo Hx1_Neo_GAPDH (63). Aproximadamente 400 pares de base correspondentes a região 5’ não traduzida (5’UTR) e ao início da região codificadora da TcRBP99 foram amplificados a partir do DNA genômico do clone CL Brener usando os iniciadores For5’KO_TcRBP99_HindIII e Rev5’KO_TcRBP99_SacI (Tabela 1) gerando o fragmento 5’KO:TcRBP99, que foi então clonado no vetor de clonagem TOPO2.1. Simultaneamente o plasmídeo Hx1_Neo_GAPDH foi digerido com as enzimas de restrição *Hind*III e *Sac*I para que em seguida o fragmento 5’KO:TcRBP99 fosse ligado à região 5’ do plasmídeo Hx1_Neo_GAPDH dando origem ao plasmídeo

5’KO:TcRBP99_Hx1_Neo_GAPDH. Da mesma maneira, aproximadamente 400 pares de base correspondentes ao final da região codificadora da TcRBP99 e da região 3’ não traduzida (3’UTR) foram amplificados do DNA genômico do clone CL Brener utilizando os iniciadores For3’KO_TcRBP99_XhoI e Rev3’KO_TcRBP99_XbaI (Tabela 1) gerando o fragmento 3’KO:TcRBP99 que em seguida foi clonado no vetor TOPO2.1. Este fragmento foi retirado do TOPO através de reação de digestão, usando as enzimas de restrição *Xba*I e *Xho*I, e em seguida clonado na região 3’ do plasmídeo 5’KO:TcRBP99_Hx1_Neo_GAPDH, previamente digerido com *Xba*I e *Xho*I, dando origem ao plasmídeo 5’KO:TcRBP99_Hx1_Neo_GAPDH_3’KO:TcRBP99 (pKO_TcRBP99_Neo). Esse plasmídeo foi utilizado como molde em reação de PCR, juntamente com os iniciadores For5’KO_TcRBP99_HindIII e Rev3’KO_TcRBP99_XbaI, para amplificação da sequencia doadora utilizada na transfecção de parasitos selvagens para interromper um dos alelos da TcRBP99.

Para deleção do segundo alelo, o plasmídeo 5’KO:TcRBP99_Hx1_Neo_GAPDH_3’KO:TcRBP99 foi digerido com as enzimas de restrição *Spe*I e *Not*I e em seguida clonado no plasmídeo Hx1_Higro_GAPDH, previamente digerido com essas mesmas enzimas. O plasmídeo 5’KO:TcRBP99_Hx1_Higro_GAPDH_3’KO:TcRBP99 (pKO_TcRBP99_Higro), gerado pela clonagem, foi usado como molde em reação de PCR juntamente com os iniciadores For5’KO_TcRBP99_HindIII e Rev3’KO_TcRBP99_XbaI, gerando a sequencia doadora com resistência a higromicina, que foi então usada para transfectar parasitos com um dos alelos deletados.

3.9.3 Vetor para re-expressar a TcRBP99

O plasmídeo pROCK com resistência a puromicina e sequencia da região codificadora da TcRBP99 foi utilizado para a re-expressar o gene desta RBP. Para isso o plasmídeo pROCK_TcRBP99::HA_Neo foi digerido com as enzimas *Xho*I e *Nhe*I liberando o fragmento contendo as sequencias referentes a região intergênica de GAPDH, o gene de resistência a neomicina e a região 3'UTR de GAPDH. Simultaneamente o plasmídeo pROCK_PAC_GFP foi digerido com essas mesmas enzimas para a liberação do fragmento contendo as mesmas sequencias de GAPDH porém com a sequencia correspondente ao gene de puromicina entre as regiões de GAPDH. Este fragmente carregando a sequencia de resistência a puromicina foi então ligado ao restante do pROCK_TcRBP99::HA_Neo dando origem ao plasmídeo pROCK_TcRBP99::HA_PAC que em seguida foi linearizado utilizando a enzima de restrição *Not*I. Parasitos nocautes para o gene da TcRBP99 foram transfectados com o plasmídeo pROCK_TcRBP99::HA_PAC que integra no locus de β-tubulina.

3.9.4 Vetor para super expressar a TcRBP300

Assim como para a TcRBP99 a super expressão do gene TcCLB.507611.300 (TcRBP300) foi realizada utilizando o plasmídeo pROCK (61). A região codificadora da TcRBP300 foi amplificada a partir do DNA genômico do clone CL-14 utilizando os iniciadores For_TcRBP300_NheI e Rev_TcRBP300::HA_XhoI, descritos na tabela 1, com sequencia para adição da *tag* de HA na porção C-terminal. Após a clonagem no vetor TOPO2.1 e digestão com as enzimas de restrição *Nhe*I e *Xho*I, o produto amplificado foi clonado no plasmídeo pROCK_GFP_Neo, previamente digerido com *Xba*I e *Xho*I, dando origem ao plasmídeo pROCK_TcRBP300::HA_Neo. Após ser linearizado com a enzima de restrição *Not*I, este plasmídeo foi usado para transfectar formas epimastigotas do clone CL Brener.

3.9.5 Vetor para nocaute do gene que codifica a TcRBP300

Nocaute da TcRBP300 foi realizado usando a tecnologia da CRISPR/Cas e duas sequencias doadoras foram usados para seleção de linhagens nocautes.

Uma das sequencias doadoras foi gerada utilizando o vetor Hx1_Neo_GAPDH. Primeiramente 400 pares de base, correspondentes ao final da região codificadora da TcRBP300 e parte da região 3'UTR, foram amplificados a partir do DNA genômico do clone CL-14 com os iniciadores For3'KO_TcRBP300_XhoI e Rev3'KO_TcRBP300_XbaI (Tabela 1). O fragmento gerado, 3'KO:TcRBP300, foi clonado no vetor TOPO2.1 e posteriormente clonado no plasmídeo Hx1_Neo_GAPDH previamente digerido com as enzimas *Xho*I e *Xba*I dando origem ao plasmídeo Hx1_Neo_GAPDH_3'KO:TcRBP300. Usando os iniciadores For5'KO_TcRBP300_KpnI e Rev5'KO_TcRBP300_SacI, fragmento de aproximadamente 340 pares de base (5'KO:TcRBP300) foi amplificado do DNA genômico do clone CL-14. Após clonagem no TOPO2.1, este fragmento foi clonado no plasmídeo Hx1_Neo_GAPDH_3'KO:TcRBP300 previamente digerido com as enzimas *Kpn*I e *Sac*I gerando o plasmídeo 5'KO:TcRBP300_Hx1_Neo_GAPDH_3'KO:TcRBP300 (pKO_TcRBP300_Neo) que foi usado como molde em reação de PCR juntamente com os iniciadores For5'_TcRBP300_KpnI e Rev3'_TcRBP300_XbaI, resultando na sequencia doadora usada na transfecção de parasitos CL-14 selvagens.

A outra sequencia doadora é constituída por parte da sequencia da TcRBP300 contendo códon de parada de tradução, em três diferentes janelas de leitura, seguido da sequência da enzima de restrição *Bam*HI.

3.10 Obtenção do sgRNA

A identificação do sgRNA foi realizada por meio do software EuPaGDT (Eukaryotic Pathogen CRISPR guide RNA Design Tool; disponível em: http://grna.ctegd.uga.edu/) no qual foi utilizado a opção SaCas9 (21-bp target sequence preceding an NNGRRT PAM site) e fornecida a sequencia da região codificadora da TcRBP300. Par de iniciadores For_sgRNA_TcRBP300 e Rev_sgRNA (Tabela 1) e um plasmídeo contendo a sequencia do *scaffold* da Cas9 de *Staphylococcus aureus* como molde geraram um produto de PCR que foi submetido à reação de transcrição *in vitro* pela T7 RNA polimerase utilizando o kit MEGAscript T7 Transcription (Thermo Fisher Scientific) de acordo com as instruções do fabricante.

3.11 Transfecção de parasitos

Formas epimastigotas do *T. cruzi* foram transfectadas utilizando o nucleofector. Para isso $4x10^7$ parasitos foram centrifugados a 1.200xg por 10 minutos e ressuspendidos em tampão de eletroporação Tb-BSF buffer (73), seguido pela adição de 20µg do plasmídeo ou sequencia doadora de interesse. Essa mistura foi então transferida para cuvetas de 0,2mm (Bio-Rad) e submetidas ao programa U033 no Amaxa Nucleofector II Device. Após a transfecção os parasitos foram transferidos para garrafas de cultura contendo 5mL de meio LIT suplementado com 10% de soro fetal bovino e vinte e quatro horas após a transfecção 200 µg/mL do antibiótico de seleção foi adicionado as culturas.

Transfecção com Cas9 recombinante, produzida por nosso grupo, foi realizada utilizando $4x10^6$ formas epimastigotas do clone CL-14 que foram centrifugados a 1.200xg por 10 minutos e ressuspendidos em tampão de eletroporação Tb-BSF buffer (73), seguido pela adição de 20µg da sequencia doadora de interesse, 15µg de Cas9 recombinante de *S. aureus* (SaCas9) e 20µg de sgRNA. Essa mistura foi então

transferida para cuvetas de 0,2mm (Bio-Rad) e submetidas ao programa U033 no Amaxa Nucleofector II Device. Após a transfecção os parasitos foram transferidos para garrafas de cultura contendo 5mL de meio LIT suplementado com 10% de soro fetal bovino

Clones foram gerados por diluição limitante após a seleção da população. Para isso os parasitos foram diluídos em meio LIT contendo 200µg/mL de antibiótico de seleção e transferidos para placas de 96 poços de maneira que fosse adicionado 0,5 parasito por poço.

3.12 Curva de crescimento

Crescimento de formas epimastigotas foi avaliado diluindo os parasitos na densidade de $2x10^6$ parasitos/mL de cultura com acompanhamento do crescimento a cada dois dias através da contagem em câmera de Neubauer.

3.13 Metaciclogênese

Ensaios de metaciclogênese foram realizados com formas epimastigotas, cultivadas em meio LIT, em fase logarítmica de crescimento que foram transferidas para um meio pobre em nutrientes. O equivalente a $2x10^7$ parasitos foi centrifugado a 1.200xg durante 10 minutos, o *pellet* foi ressuspendido em 10mL de meio RPMI (Gibco), sem soro fetal bovino, e as garrafas foram mantidas deitadas sem agitação em estufa a 28ºC (74). Após 4 dias, a cada 2 dias lâminas com aproximadamente 20µl do sobrenadante foram preparadas e coradas com giemsa. Porcentagem de tripomastigotas metacíclicas foi estimada pela contagem em microscópio com base na morfologia.

3.14 Imunofluorescência

Localização celular das proteínas foi analisada por meio de imunofluorescência. Para isso formas epimastigotas em fase logarítmica de crescimento foram

centrifugadas a 1.200xg durante 10 minutos, o *pellet* foi ressuspendido em solução salina 1X (PBS) e centrifugado novamente nas mesmas condições. Em seguida o *pellet* foi ressuspendido em aproximadamente 200µl de paraformaldeído (4%) por 5 minutos a temperatura ambiente para fixação. Após a fixação os parasitos foram centrifugados novamente, o *pellet* foi ressuspendido em PBS e os parasitos foram colocados em lâmina de vidro previamente tratada com poli-lisina. Os parasitos foram permeabilizados durante 10 minutos com 0,1% de Triton X-100, em sequência foram lavados três vezes com PBS 1X e então adicionada a solução de bloqueio (PBS 1% e Tween 0,2% diluído em PBS) por uma hora a temperatura ambiente. Em seguida os parasitos foram incubados durante 40 minutos com anticorpo anti-HA (Roche) na concentração de 1:50, diluído em solução de bloqueio. Após três lavagens com solução de bloqueio os parasitos foram incubados com anticorpo secundário anti rato conjugado com Alexa 488 (Molecular Probes/ Life Technologies) e após 30 minutos as lâminas foram lavadas três vezes com solução de bloqueio e três vezes com PBS 1X. Por fim as lâminas foram montadas com *prolong gold antifade* contendo DAPI (Thermo Fisher Scientific) e lamínulas foram seladas com esmalte. As lâminas foram observadas no Microscópio Óptico de Fluorescência Axio Imager Z2 – ApoTome 2-Zeiss no Centro de Microscopia da Universidade Federal de Minas Gerais.

3.15 Imunoprecipitação

Extrato proteico total de epimastigotas de *T. cruzi* expressando a TcRBP99::HA foi lisado a 4°C em tampão de lise (Tris-HCl pH7.5 10mM; NaCl 140mM; NP-40 1%). Em seguida, o lisado foi centrifugado a 1.200xg por 10 minutos, o sobrenadante foi adicionado à *bead* EZview Red Anti-HA (Sigma – Aldrich) e incubado à 4°C por cerca de 16 horas. As *beads* foram então centrifugadas a 8.200xg por 30 segundos e o sobrenadante, correspondente às proteínas que não se ligaram, foi transferido para

um novo tubo para análises posteriores. Em seguida, as *beads* foram lavadas com tampão de lise e as proteínas que se ligaram foram eluidas em tampão de eluição (62,5 mM Tris-HCl pH 6,8; 10% de glicerol; 2% de SDS; 5% de β-mercaptoetanol; 0,002% de azul de bromofenol) para posterior análise por western blot. Parasitos selvagens foram usados como controle negativo. RNA total foi extraído de alíquotas do lisado inicial, do sobrenadante e do eluato.

3.16 Extração de RNA, síntese de cDNA e Real time PCR

Extração de RNA total das frações correspondentes ao lisado inicial, ao sobrenadante e ao eluato da imunoprecipitação foi realizada com trizol. Para isso trizol foi adicionado a essas frações que então foram incubadas no gelo por 5 minutos. Em seguida 200µL de clorofórmio foi adicionado ao tubo que foi agitado e deixado no gelo por 3 minutos. Após centrifugação de 12.400xg por 15 minutos à 4ºC, o sobrenadante foi transferido para um novo tubo, seguido pela adição de 500µL de isopropanol, incubação no gelo por 10 minutos e centrifugação a 12.400xg por 10 minutos à 4ºC. Após o sobrenadante ser descartado, 1mL de etanol 70% foi adicionado, sem ressuspender o *pellet*, e o tubo foi centrifugado novamente a 7.500xg por 5 minutos à 4°C.

Aproximadamente $1x10^8$ formas epimastigotas, em fase logarítmica de crescimento, foram usadas para a extração total de RNA utilizado em ensaios de PCR em tempo real (RT-PCR) e para o sequenciamento de mRNA. Neste caso a extração foi realizada com o kit Illustra RNAspin Mini (GE Healthcare) seguindo as instruções do fabricante.

cDNA fita simples foi sintetizado com a Transcriptase Reversa SuperScript II Reverse Transcriptase kit first-strand synthesis (Invitrogen) e $Oligo(dT)_{18}$ de acordo com as instruções do fabricante. Para as reações de RT-PCR foram usados 0,5µM de

iniciadores senso e anti-senso específicos, descritos na tabela 1, e SYBR Green 1X (Bio-Rad). As reações foram submetidas ao equipamento Applied Biosystems 7900HT Fast Real-Time PCR System (Life Technologies) e analisadas utilizando o software SDS disponibilizado pela Thermo Fisher Scientific. Níveis relativos de mRNA foram normalizados com iniciadores específicos para o gene codificante para a proteína ribosomal L9, RpL9 (TcCLB.504181.10), descrita na tabela 1, e quantificados usando o método de curva padrão baseado no ciclo ($2^{-\Delta\Delta ct}$) (75), sendo o valor de *threshold* definido em 0,3792535.

3.17 RNA-Seq

RNA total foi extraído de formas epimastigotas de parasitos selvagens, em triplicata, e nocautes (dois clones em duplicata) para a TcRBP99 como descrito no item anterior. As amostras foram enviadas ao Instituto René Rachou (Fiocruz Minas) para análise da qualidade, realizada por bioanalyzer 2100 (Agilent), quantificação por RT-PCR, usando o kit KAPA library quantification (KAPA Biosystems), preparo da biblioteca com o kit v2 Illumina® TruSeq® RNA Sample Preparation (Illumina) e sequenciamento de nova geração na plataforma MiSeq. Análise das *reads*, mapeamento e análise da expressão diferencial foram realizadas pela aluna de mestrado Wanessa Moreira Goes. Em resumo a qualidade das *reads* foi avaliada com a ferramenta FastQC (disponível em: projects/fastqc/) e “trimada” com o software Trimmomatic (76). Em seguida as sequencias foram mapeadas no genoma de CL Brener do *T. cruzi*, versão 9.0, disponível no banco de dados TritrypDB, usando o software TopHat2 (77) permitindo no máximo 2 mismatches e que uma mesma sequencia alinhasse no máximo em 20 regiões. O software HTSeq (78) foi usado para determinar o número de *reads* alinhadas nas regiões codificadores e o pacote DESeq2 (disponível em: https://bioconductor.org/packages/release/bioc/html/DESeq2.html) foi

utilizado para determinar os genes diferencialmente expressos, isto é, transcritos cujos níveis apresentam diferenças com valores de log2 de *fold change* maior ou menor que 1 e *p-value* ajustado menor que 0,05.

3.18 Extrato de formas epimastigotas do *T. cruzi*

Formas epimastigotas em fase logarítmica de crescimento foram centrifugadas a 1.200xg por 10 minutos, ressuspendido em PBS 1X e centrifugados novamente nas mesmas condições. O *pellet* foi ressuspendido em tampão de lise gelado que contêm: tampão RIPA 1X (50 mM Tris-HCl pH8.0, 150mM NaCl, 1% NP-40, 0.5% deoxicolato de sódio, 0,1% SDS) e tampão de amostra 4X (62,5 mM Tris-HCl pH 6,8; 10% de glicerol; 2% de SDS; 5% de β-mercaptoetanol; 0,002% de azul de bromofenol). Em seguida os extratos foram aquecidos a 100°C durante 10 minutos.

3.19 Western Blot

Extrato de parasitos ou frações coletadas na imunoprecipitação foram separados em gel de poliacrilamida na concentração 12,5% por aproximadamente 1 hora a 180 volts em tampão de corrida. Em seguida foram transferidos para membrana de nitrocelulose (Amersham Protran 0.45 µm - GE Healthcare) a 100 volts durante 2 horas em tampão de transferência (25 mM Tris, 192 mM glicina, 20% (v/v) metanol). Após a transferência a membrana foi corada com Ponceau (0,1% de Ponceau diluído em 3% de ácido acético), incubada com solução de bloqueio (5% de leite em pó desnatado dissolvido em PBS 1X contendo 0,1% de tween 20) por no mínimo uma hora e em seguida incubada com o anticorpo anti-HA (Sigma-Aldrich), diluído 1:5.000 em solução de bloqueio, por aproximadamente 16 horas. A membrana foi então lavada três vezes com PBS 1X contendo 0,1% de tween 20, incubada com anticorpo secundário anti-mouse (Sigma-Aldrich), diluído 1:10.000 em solução de bloqueio, por aproximadamente 1 hora, lavada três vezes com PBS 1X com 0,1% de tween 20 e

uma vez com PBS 1X. Luminata (Millipore) foi usado para revelar a membrana, sendo as imagens capturadas no fotodocumentador (Gel LOGIC 1500 Imaging System).

3.20 Análise de sequencias 3’ UTR para identificação de motivos conservados

Mapeamento das *reads* provenientes de todas as bibliotecas usadas neste trabalho foram usadas para determinar a região 3’UTR dos genes diferencialmente expressos entre as linhagens nocautes e os parasitos selvagens. Mapeamento foi visualizado usando o programa *Integrative Genome View* (IGV) (79), e a sequencia correspondente a região 3’UTR, começando na primeira base depois do códon de parada e terminando até onde há mapeamento de *reads* antes do próximo gene, foi extraída. Usando o programa de edição de sequencia Life Science Tool (disponível em: http://www.fr33.net/), a sequencia de cDNA foi transformada em mRNA e essas sequencias foram submetidas ao programa MEME para identificação de motivos conservados (80).

4. RESULTADOS - PARTE I

4.1 Identificação das proteínas de ligação a RNA no genoma do *T. cruzi*

Considerando o fato da transcrição gênica dos tripanosomatídeos ocorrer de forma policistrônica e que as proteínas de ligação ao RNA (RBP) têm importante papel na regulação dos níveis dos mRNAs no citoplasma, seja mediando a sua estabilização, degradação ou sua taxa de tradução, novas buscas por sequencias contendo motivos de ligação ao RNA no genoma do clone CL Brener do *T. cruzi*, depositado no banco de dados TritrypDB (disponível em: http://tritrypdb.org/tritrypdb/), foram conduzidas. Para isso foram realizadas análises, utilizando o programa Blast, de sequências contendo domínios das famílias pumilio (PUF), *zinc finger*, motivo de reconhecimento ao RNA (RRM), proteína de ligação a cauda poli(A) (PABP) e alba. As sequencias que determinam esses motivos foram extraídas do banco de dados Pfam (disponível em: http://pfam.xfam.org/). Utilizando como critério de seleção resultados de alinhamento com identidade igual ou superior a 85% e *e-value* inferior a 10^{-9}, foram identificadas 253 sequencias codificadoras de proteínas codificadas no genoma do clone CL Brener que possuem motivos de ligação ao RNA (Tabela 3). É importante ressaltar que devido à natureza híbrida de CL Brener, este número total de sequencias identificadas em nossas análises representa 147 genes codificantes de proteínas, uma vez que, em muitos casos os dois alelos do mesmo gene foram anotados separadamente. Na tabela 3 observa-se que o domínio mais frequentemente presente nas RBPs identificadas foi o RRM, o qual aparece em 124 sequencias sendo que no caso da proteína de ligação à cauda poli-A, previamente caracterizada no parasito (81), além deste domínio é encontrado também o motivo PABP. A análise mostrou ainda a presença de 101, 22 e 6 sequencias que possuem os motivos *zinc finger*, pumilio e alba respectivamente (Tabela 3). A lista completa das

sequencias contendo motivos de ligação a RNA identificadas no genoma do clone CL Brener do *T. cruzi* pode ser encontrada como material suplementar do artigo recentemente publicado pelo nosso grupo e que se encontra no Apêndice V (24).

Dentre todas as sequencias identificadas, a grande maioria, que corresponde a 229 sequencias, já havia sido identificada em outros estudos, porém um número significativo de novas sequencias codificadoras de RBPs em *T. cruzi* foi encontrado. Dentre os genes codificantes para RBPs ainda não identificados, 12 sequencias (TcCLB.503757.30, TcCLB.503719.39, TcCLB.504089.60, TcCLB.504001.10, TcCLB.504089.70, TcCLB.504001.20, TcCLB.510297.10, TcCLB.509805.240, TcCLB.511267.20, TcCLB.463297.10, TcCLB.510877.30, TcCLB.510877.40) anotadas no banco de dados do genoma do *T. cruzi* como genes hipotéticos codificam, de acordo com essas análises, proteínas de ligação a RNA.

Tabela 3: Número de sequencias identificadas no genoma de CL Brener contendo domínios de ligação a RNA

Domínio de ligação a RNA	Sequencias identificadas
RRM	122
RRM + PABP	2
Zinc Finger	101
Pumilio	22
Alba	6
Total	253

4.2 Proteínas de ligação a RNA diferencialmente expressas ao longo do ciclo de vida do *T. cruzi*

Em colaboração com o professor Najib El-Sayed, da Universidade de Maryland nos EUA, foram gerados recentemente dados de RNA-Seq a partir de formas epimastigotas, cultivada em meio LIT, amastigotas intracelulares, obtidas 60 e 96 horas após a infecção de *human foreskin fibroblasts* (HFF), e de formas tripomastigotas liberadas em cultura dos clones CL Brener e CL-14 de *T. cruzi*. Com

o sequenciamento dos mRNAs destas distintas formas dos clones CL Brener e CL-14 do *T. cruzi* foi possível analisar a expressão gênica global, com o objetivo de buscar diferenças na expressão de genes que poderiam explicar as diferenças de virulência observadas no comportamento dessas cepas quando avaliadas em ensaios de infecção *in vitro* e *in vivo*. De posse desses dados de transcriptoma, como parte inicial do meu trabalho de Tese, buscamos avaliar a expressão dos transcritos correspondentes ao grupo de genes codificantes para as proteínas de ligação a RNA nas diferentes fases do ciclo de vida dos dois clones do parasita. Essas análises de expressão gênica global foram conduzidas pelo nosso grupo em colaboração com o grupo do professor Najib El-Sayed, da Universidade de Maryland nos EUA e o artigo resultante desse trabalho, publicado na revista PLos Pathogens, encontra-se no apêndice III (24).

De início foram conduzidas pesquisas para identificar os transcritos codificantes para proteínas de ligação a RNA com expressão aumenta na forma epimastigota, que habita o hospedeiro invertebrado, em comparação com as formas amastigota e tripomastigota, encontradas no hospedeiro vertebrado. Nessas análises foram utilizados apenas os dados de RNA-Seq proveniente do clone CL Brener, uma vez que não foram geradas bibliotecas com qualidade adequadas a partir de culturas de epimastigota do clone CL-14. Para essas análises foram desconsideradas também os dados referentes as formas amastigotas coletadas 96 horas após a infecção uma vez que a análise da expressão gênica global mostrou que não há grandes diferenças entre os genes expressos neste tempo de infecção e os transcritos expressos nas formas tripomastigotas liberadas em cultura. Assim, foram comparados níveis de mRNA de genes que codificam RBPs nas formas epimastigotas (formas replicativas presentes no inseto) com as formas tripomastigotas não replicativas liberadas em

cultura e amastigotas coletadas 60 horas após a infecção (formas replicativas presentes no hospedeiro mamífero) Contagens normalizadas referente às quantidades de *reads* alinhadas na região codificadora de genes de cada uma das sequencias codificantes de proteínas de ligação a RNA, revelaram a presença de cinco transcritos com expressão aumentada em epimastigotas em relação a tripomastigotas (Figura 8A, painel superior) e amastigotas com 60 horas após a infecção (Figura 8A, painel inferior). Como pode ser observado na figura 8B, essas cinco RBPs com expressão aumentada em epimastigotas, comparado a ambos os estágios presentes no hospedeiro mamífero, possuem entre 173 e 257 aminoácidos, sendo que três delas apresentam o domínio *zinc finger* e duas o domínio RRM. É interessante notar que as RBPs com domínio *zinc finger* possuem um único domínio enquanto que as proteínas de ligação a RNA com domínio RRM apresentam dois desses domínios. Ao analisar a expressão desses genes codificantes de proteínas de ligação a RNA com expressão aumentada em epimastigotas, é possível observar que os genes que contêm o domínio *zinc finger* de ligação a RNA (TcCLB.506739.99, TcCLB.506931.4 e TcCLB.510131.44) apresentam níveis significativos de expressão apenas em formas epimastigotas e níveis muito baixos nas formas presentes no hospedeiro mamífero (Figura 8C). Os outros dois genes, que possuem o motivo RRM, apresentam expressão 1,2 vezes e 1,9 vezes aumentada em epimastigota comparado com amastigota e tripomastigota respectivamente. Além disso, é possível observar que os transcritos referentes a essas proteínas (TcCLB.511727.270 e TcCLB.511727.290) também são expressos em níveis relativamente elevados nos estágios que habitam o hospedeiro vertebrado (Figura 8C). Com base nessa análise o gene TcCLB.506739.99 foi escolhido para caracterização uma vez que trata-se de um gene cujos transcritos apresentaram grande diferença de expressão

(aproximadamente 8 vezes) quando epimastigotas, amastigota e tripomastigotas são comparadas. O gene codificante para esta RBP foi denominado TcRBP99.

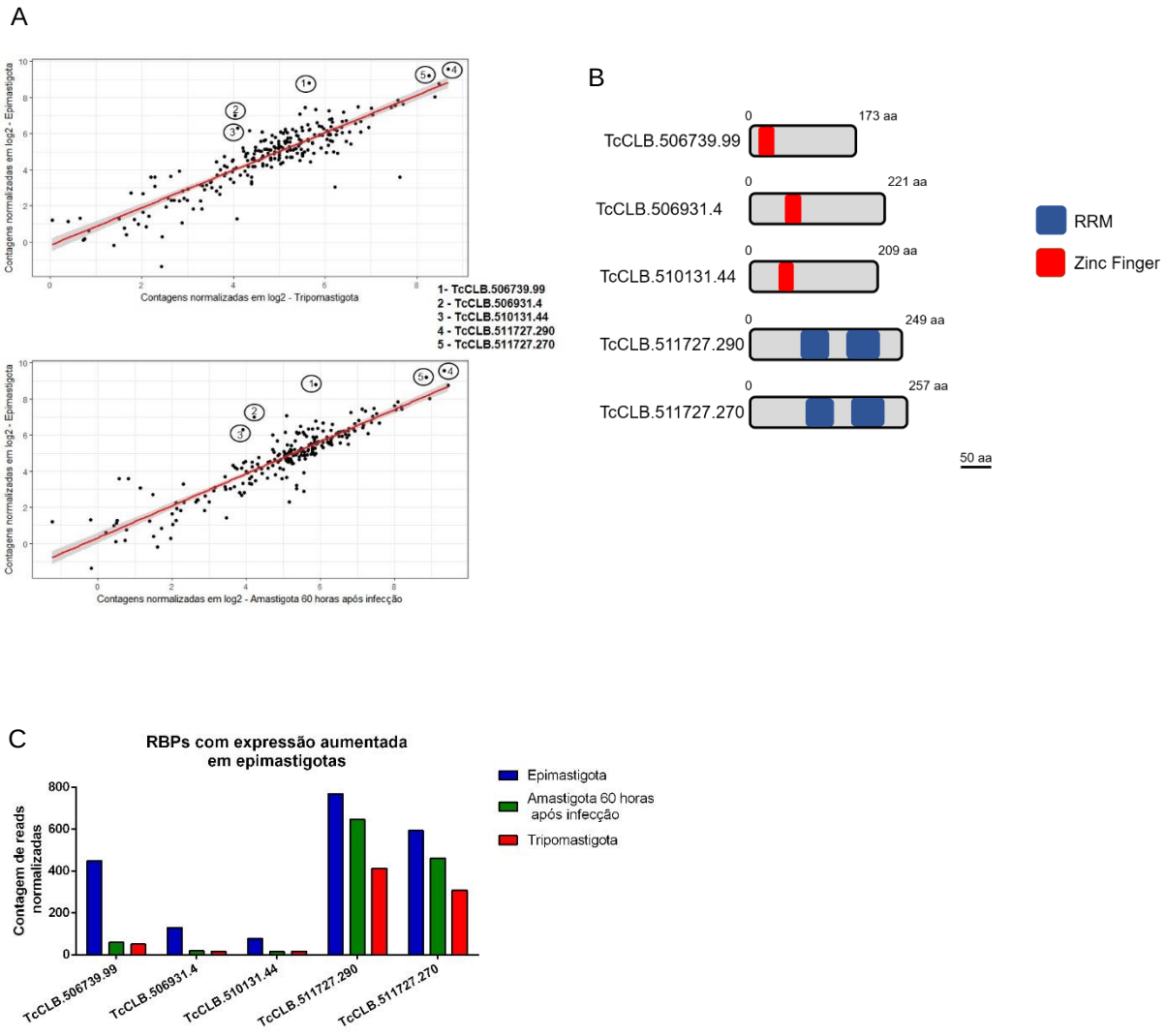


Figura 8: Proteínas de ligação a RNA com expressão aumentada em epimastigotas. (A) A partir de dados de RNA-Seq gerados com mRNAs provenientes de formas epimastigotas, amastigotas coletadas 60 horas após a infecção *in vitro* e tripomastigotas liberadas em culturas de células infectadas, 5 genes codificadores de proteínas de ligação a RNA com expressão aumentada em epimastigotas foram identificados. Scatterplots comparando os níveis dos transcritos codificantes para RBPs entre as formas epimastigota e tripomastigota (painel superior) e epimastigota e amastigotas (painel inferior). Os círculos indicam genes de RBPs com expressão aumentada em epimastigotas em ambas as comparações. (B) Representação esquemática das sequencias de aminoácidos das 5 RBPs com expressão aumentada em epimastigotas, indicando a posição dos domínios *zinc finger* (em vermelho) e RRM (azul). (C) Contagens normalizadas das *reads* que alinharam nas sequencias dos cinco genes codificadores de RBPs com expressão aumentada em formas epimastigotas. Em azul estão representadas as contagens a partir das bibliotecas de formas epimastigotas, em verde amastigotas coletadas 60 horas após a infecção e em vermelho formas tripomastigotas.

4.3 Análises *in silico* do gene codificador da TcRBP99

Como mostra a figura 9A, o mapeamento das *reads* geradas pelo RNA-Seq correspondentes ao gene da TcRBP99 são mais numerosas em formas epimastigotas (retratadas pela cor azul) quando comparadas com as formas amastigota, em verde, e tripomastigota, em vermelho. A partir desse mapeamento, o tamanho das regiões 5'UTR (iniciando na base anterior à adenina do códon da metionina inicial e terminando no final do alinhamento das *reads*) e 3'UTR (corresponde a região compreendida entre a base posterior à última base do códon de parada até o final do alinhamento das *reads*) foram estimadas em aproximadamente 312 e 767 nucleotídeos respectivamente. Desas forma, pode-se estimar que o mRNA mensageiro do gene TcCLB.506739.99 possui entre 1.700 e 1.800 nucleotídeos, correspondente à soma de 312 nucleotideos da região 5' UTR, 522 da região codificadora, 767 da região 3'UTR e entre 100 a 200 nucleotídeos de cauda poli-A. As análises confirmaram os dados depositados no banco de dados TritrypDB que mostram que a região codificadora do gene codificante para a proteína de ligação a RNA TcRBP99, anotada como "*zinc finger (CCCH type) protein, putative*", localizado no cromossomo 3 do alelo Esmeraldo-like do clone CL Brener do *T. cruzi*, é composta por 522 pares de bases, codificando, portanto, uma proteína com 173 aminoácidos. Além da presença do domínio *zinc finger*, localizado entre os nucleotídeos 15 e 41, o gene codificante para a TcRBP99 apresenta ponto isoelétrico (pI) de 8,49 que indica um caráter básico da proteína que é característica de RBPs pois favorecem a interação entre a proteína e as cargas negativas do RNA. Como pode ser observado na tabela 4 que mostra a comparação entre a proteína de ligação a RNA codificada pelo alelo esmeraldo TcCLB.506739.99 (TcRBP99) e seu alelo não-esmeraldo (TcCLB.510819.119), as quais apresentam 97,7% de similaridade na sequencia de

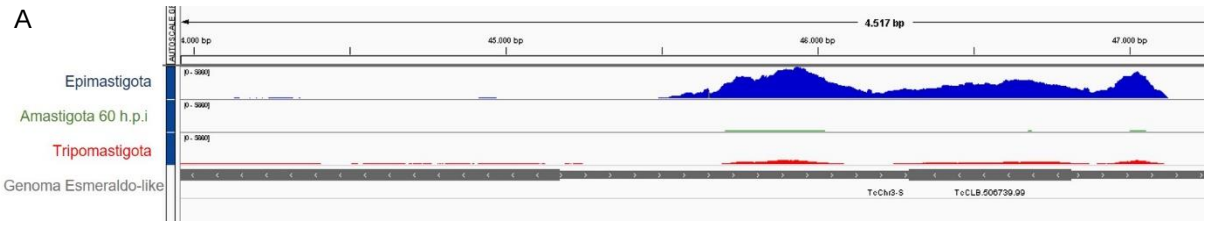
aminoácidos, os dois alelos apresentam apenas pequenas diferenças na posição referente a localização no cromossomo 3 e no ponto isoelétrico das proteínas por eles codificadas.

Tabela 4: Principais características dos alelos codificantes para a proteína de ligação a RNA com expressão aumenta em epimastigotas.

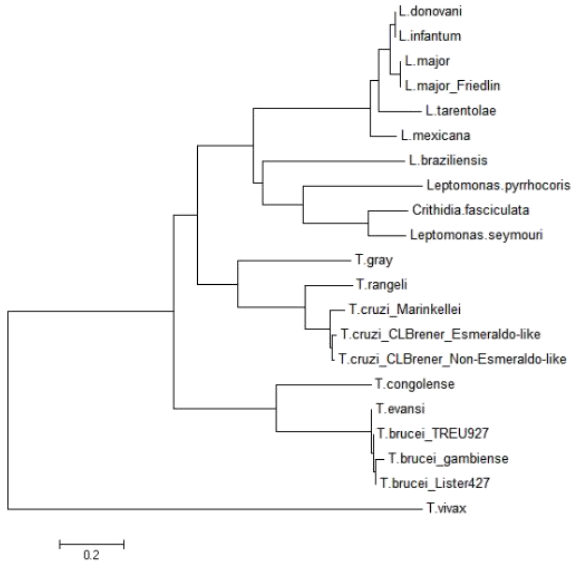
	TcCLB.506739.99	**TcCLB.510819.119**
Alelo	Esmeraldo-like	Não Esmeraldo-like
Localização	Posição 46.294 a 46.815 do cromossomo 3	Posição 46.651 a 47.172 do cromossomo 3
Tamanho do gene	522 bases	522 bases
Tamanho da proteína	173 aminoáciods	173 aminoáciods
Localização do motivo *zinc finger*	Entre as bases 15 e 41	Entre as bases 15 e 41
Ponto isoelétrico	8,49	8,29

Sequencias de genes sintênicos à TcRBP99 foram extraídas do banco de dados TritrypDB que contêm além de dados de genoma do *T. cruzi* sequencias de vários outros tripanosomatídeos. A análise de filogenia, utilizando o método *Neighbor joining*, das sequencias ortólogas da TcRBP99 agrupou as sequencias em 4 principais grupos: *T. cruzi*, *T. brucei*, Leptomonas e Leishmania (Figura 9B). O alinhamento múltiplo revelou alta conservação na região N-terminal bem como do motivo *zinc finger* (C-X8-C-X5-C-X3-H), localizado nesta região da proteína, entre as sequencias ortólogas de todos os organismos analisados (Figura 10B). Não foi encontrada homologia com outros eucariotos.

A


4.517 bp
44.000 bp
45.000 bp
46.000 bp
47.000 bp
Epimastigota
Amastigota 60 h.p.i
Tripomastigota
Genoma Esmeraldo-like
TcChr3-S
TcCLB.506739.99

B


L.donovani
L.infantum
L.major
L.major_Friedlin
L.tarentolae
L.mexicana
L.braziliensis
Leptomonas.pyrrhocoris
Crithidia.fasciculata
Leptomonas.seymouri
T.gray
T.rangeli
T.cruzi_Marinkellei
T.cruzi_CLBrener_Esmeraldo-like
T.cruzi_CLBrener_Non-Esmeraldo-like
T.congolense
T.evansi
T.brucei_TREU927
T.brucei_gambiense
T.brucei_Lister427
T.vivax
0.2

C

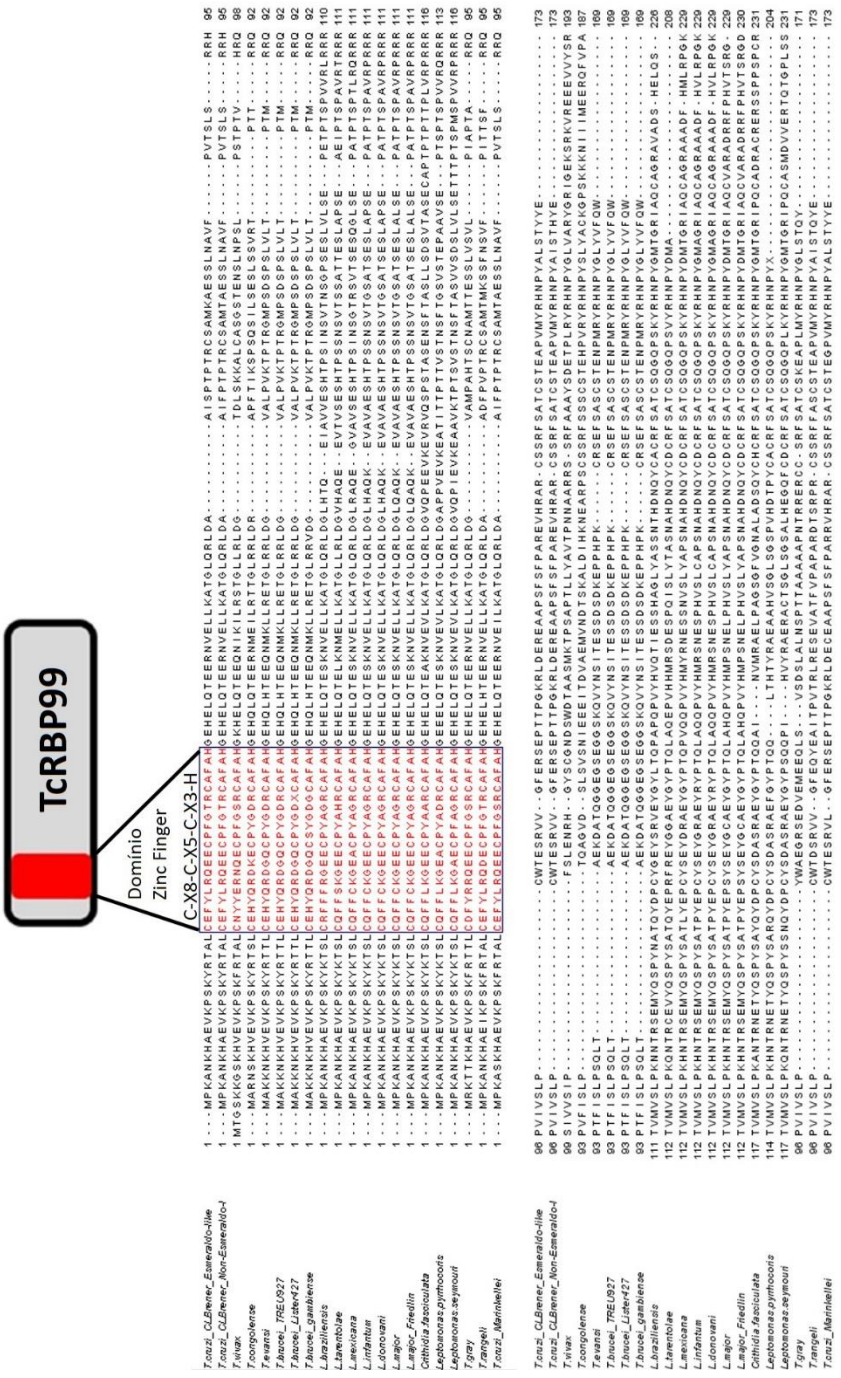


Figura 9: Análises *in sílico* da TcRBP99. (A) Mapeamento da *reads* geradas pelo RNA-Seq a partir de mRNAs de epimastigotas (azul), amastigotas coletadas 60 horas após a infecção (verde) e tripomastigotas (vermelho) no cromossomo 3 do genoma Esmeraldo-like do clone CL Brener do *T. cruzi*. As barras cinzas mais grossas indicam região codificadora e as mais finas as regiões intergênicas. (B) Árvore filogenética construída com os genes sintênicos, usando o software MEGA. (C) Alinhamento múltiplo dos genes sintênicos a TcRBP99 depositados no banco de dados TritrypDB. Os genes sintênico extraídos do banco de dados TritrypDB são representados pelos identificadores: TcCLB.506739.99 (*T. cruzi* Esmeraldo), TcCLB.510819.119 (*T. cruzi* Não Esmeraldo), TvY486_0501060 (*T. vivax*), TcIL3000_5_1490 (*T. congolense*), TevSTIB805.5.1760 (*T. evansi*), Tb927.5.1570 (*T. brucei* 927), Tb427.05.1570 (*T. brucei* 427), Tbg972.5.2150 (*T. brucei gambiense*), LbrM.15.0140 (*L. braziliensis*), LtaP15.0130 (*L. tarentolae*), LmxM.15.0140 (*L. mexicana*), LinJ.15.0140 (*L. infantum*), LdBPK_150140.1 (*L. danovani*), LMJLV39_150006500 (*L. major*), LmjF.15.0140 (*L. major* Friedlin), CFAC1_060021400 (*C. fasciculata*), LpyrH10_38_0260 (*L. pyrrhocoris*), Lsey_0162_0120 (*L. seymouri*), DQ04_00711020 (*T. grayi*), TRSC58_03397 (*T. rangeli*).

4.4 Geração de linhagens nocautes para a TcRBP99

Linhagens nocautes para a TcRBP99 foram geradas por recombinação homóloga utilizando plasmídeos contendo o gene de resistência à neomicina (ou G418) (Hx1_Neo_GAPDH) para interrupção do primeiro alelo e com resistência à higromicina (Hx1_Higro_GAPDH) para interrupção do segundo alelo. Foram adicionados a esses plasmídeos partes correspondentes às regiões 5’ e 3’ do gene da TcRBP99, amplificadas a partir de DNA genômico do clone CL Brener, para serem utilizadas no processo de recombinação homóloga após a transfecção (Figura 10A a 10C). Epimastigotas do clone CL Brener foram transfectadas inicialmente com o fragmento linear contendo a sequencia do gene de resistência à neomicina flanqueadas por sequencias do gene da TcRBP99 e tanto a população quanto os clones foram selecionados com a droga G418 (Figura 10D e 10E). Ensaios de PCR convencional utilizando diferentes conjuntos de iniciadores indicaram a integração da sequencia contendo o gene de resistência a neomicina no *locus* da TcRBP99 como pode ser observado na figura 10F. Ensaios de PCR em tempo real mostraram redução de 50% nos níveis de expressão de mRNAs codificantes para a TcRBP99, indicando que houve interrupção de um dos alelos desta RBP nos clones 3 e 7 (Figura 10G).

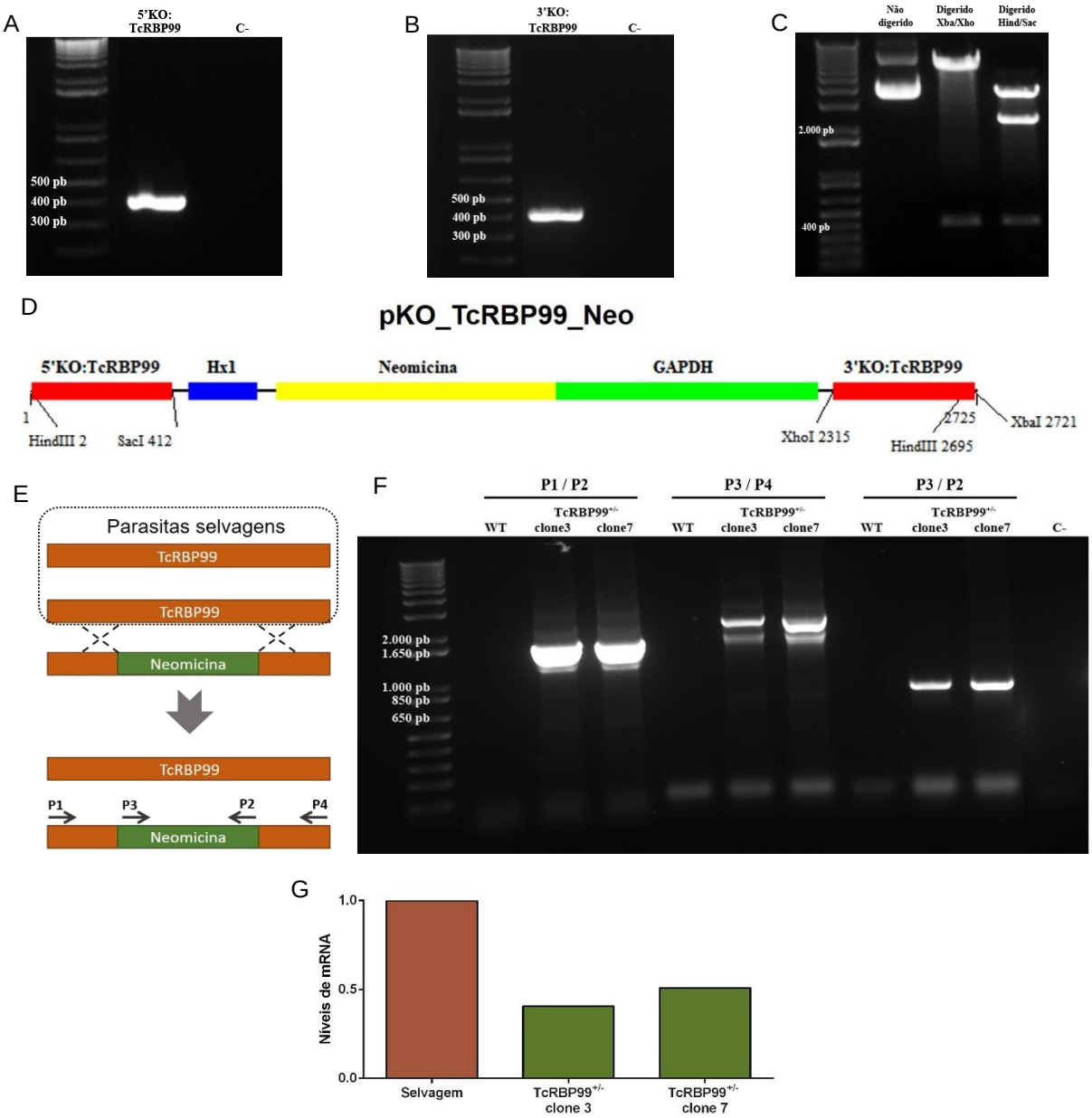


Figura 10: Construção do vetor para interrupção do primeiro alelo do gene codificante para a TcRBP99 e obtenção de linhagens hemi-nocautes. Sequencias de 400 pares de base corresponde a (A) porção final da região 5'UTR e início da região codificadora da TcRBP99 e (B) final da região codificadora desta mesma RBP e início da porção 3'UTR foram amplificadas a partir do DNA genômico do clone CL Brener. (C) Confirmação da construção do plasmídeo, contendo a sequencia doadora, através da digestão com as enzimas de restrição *Xba*I e *Xho*I (fragmento de 400 pares de base) e *Hind*III e *Sac*I (fragmentos de 2.283 e 400 pares de base). (D) Sequencia final contendo as sequencias referentes a 5'UTR de Hx1 (azul), a região codificadora do gene de neomicina (amarelo) e a região 3'UTR de GAPDH (verde), flanqueadas por sequencias homologas a TcRBP99 (vermelho) usada para interrupção de um dos alelos da TcRBP99 por recombinação homóloga. (E) Representação esquemática da interrupção de um dos alelos pela construção contendo o gene de resistência a neomicina. Setas indicam o local de anelamento dos iniciadores usados para confirmar a integração da construção. (F) DNA genômico de parasitos selvagens, da população e de dois clones foram usados para confirmar a integração do plasmídeo contendo sequencia de neomicina no *locus* da TcRBP99. Os conjuntos de iniciadores P1/P2 e P3/P4 anelam na TcRBP99 e no gene de neomicina, mostrando a integração no *locus* da TcRBP99, enquanto o conjunto P3 e P2 amplificam o gene de neomicina. (G) PCR em tempo real com iniciadores específicos para a TcRBP99.

Considerando que os dois alelos codificantes para a TcRBP99 são 98,7% similares na sequencia de nucleotídeos (Figura 11A), as mesmas sequencias da TcRBP99 presentes no plasmídeo pKO_TcRBP99_Neo foram usadas para interromper o segundo alelo da TcRBP99 por recombinação homóloga. A construção do plasmídeo para a interrupção do segundo alelo foi realizada por meio da digestão da construção contendo neomicina, de maneira que houvesse a liberação das sequencias referentes a Hx1, neomicina e GAPDH, para que o restante do plasmídeo fosse clonado no plasmídeo Hx1_Higro_GAPDH, previamente digerido, gerando a sequencia pKO_TcRBP99_Higro (Figura 12A). Epimastigotas do clone 3, obtido da população de parasitos resistentes a G418 transfectados com a construção para interrupção do primeiro alelo, foram submetidas à transfecção com o plasmídeo pKO_TcRBP99_Higro e tanto a população quanto os clones foram selecionados com as drogas G418 e higromicina (Figura 12B). Para verificar a integração da sequencia doadora no segundo alelo da TcRBP99, diferentes conjuntos de iniciadores foram utilizados. PCR com iniciadores utilizados para verificar a integração da construção contendo neomicina revelou que houve a integração desta construção no *locus* da TcRBP99, tanto no clone hemi-nocaute quanto nos clones nocautes (Figura12C). PCR realizada com iniciadores para verificar se a construção com higromicina integrou no segundo alelo da TcRBP99, resultou em uma banda de tamanho esperado apenas nos clones resistente as duas drogas, indicando que o segundo alelo foi interrompido pela construção (Figura 12D) e, que, portanto, foram gerados parasitos nocautes. Por fim, amplificação da região codificadora no DNA genômico de parasitos nocautes revelou a ausência da banda correspondente ao gene endógeno da TcRBP99 que contêm 522 pares de base. Considerando que os dois alelos da TcRBP99 tenham sido interrompidos nos clones nocautes, a amplificação com esses iniciadores resulta

em dois fragmentos de 2.363 e 2.515 pares de base, os quais, por terem tamanhos similares, apresentam padrão de migração no gel como uma única banda. Como esperado, parasitos hemi-nocautes, ou seja, nos quais apenas um dos alelos foi interrompido, apresentam uma banda de tamanho correspondente ao gene endógeno, 522 pares de bases, e outra banda maior que corresponde ao gene interrompido por Hx1_Neo_GAPDH, fragmento de 2.363 pares de base (Figura 12E). Análise dos níveis de expressão de mRNAs correspondentes ao gene codificante para a TcRBP99, por PCR em tempo real, confirmou a obtenção do nocaute do gene codificantes para a TcRBP99 uma vez que a ausência de expressão de mRNAs nos clones 1 e 2 foi observada (Figura 12F).

Figura 11: Alinhamento dos alelos da TcRBP99. Alinhamento entre as sequencias de nucleotídeos dos alelos Esmeraldo (TcCLB.506739.99) e não Esmeraldo (TcCLB.510819.119) da TcRBP99. As sequencias foram extraídos do banco de dados TritrypDB e alinhadas usando o programa Multialin (disponível em http://multalin.toulouse.inra.fr/multalin/).

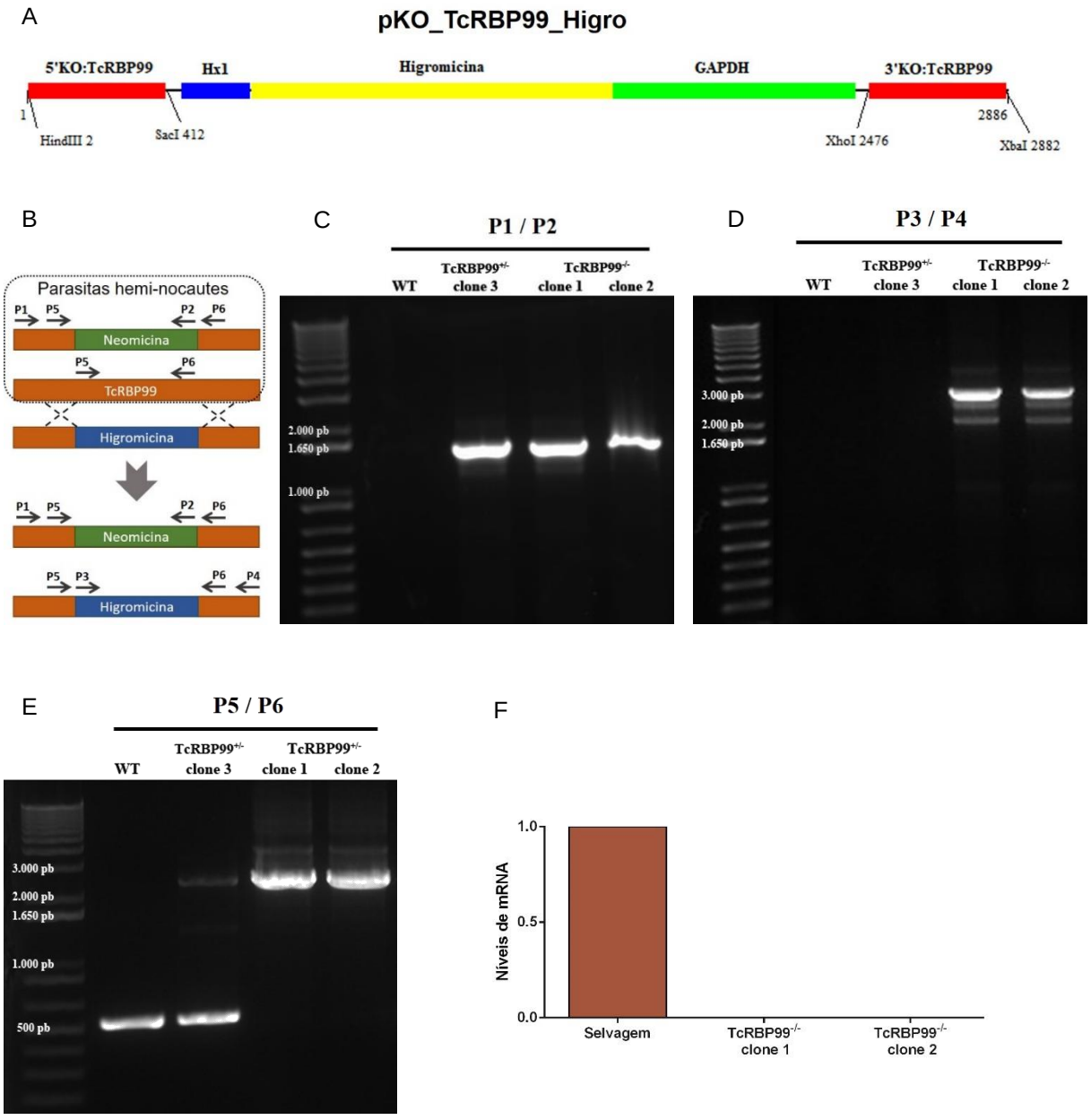


Figura 12: Construção do vetor para interrupção do segundo alelo do gene codificante para a TcRBP99 e obtenção de linhagens nocautes. Interrupção do segundo alelo da TcRBP99 foi realizada por recombinação homóloga com as mesmas sequencias usadas para a interrupção do primeiro alelo. (A) Sequencia contendo as sequencias referentes a 5'UTR de Hx1 (azul), a região codificadora do gene de higromicina (amarelo) e a região 3'UTR de GAPDH (verde), flanqueadas por sequencias homologas a TcRBP99 (vermelho). (B) Representação esquemática da interrupção do segundo alelo da TcRBP99, onde parasitos hemi-nocautes foram transfectados com a construção contendo resistência à higromicina. Setas indicam o local de anelamento dos iniciadores usados para confirmar a integração da construção. DNA genômico de parasitos selvagens, de um clone hemi-nocaute e de dois clones nocautes foram usados como molde para confirmar a integração da sequencia doadora no *locus* da TcRBP99. (C) O conjunto de iniciadores P1/P2 mostram a inserção da construção contendo neomicina no *locus* da TcRBP99 em todos os clones. Fragmento esperado de 1.648 pares de bases. (D) Os iniciadores P3/P4 indicam a presença de higromicina no *locus* da TcRBP99 nos parasitos nocautes. Fragmento esperado de 2.479 pares de bases. (E) Iniciadores para amplificar a região codificadora do gene da TcRBP99 (P5 e P6) mostram a ausência da banda correspondente ao alelo da TcRBP99 que possui 522 pares de bases. (F) PCR em tempo real com iniciadores específicos para a TcRBP99.

4.5 Geração de linhagens nocautes re-expressando o gene de TcRBP99

A geração de parasitos nocautes re-expressando a proteína de ligação a RNA TcRBP99 foi realizada, utilizando o vetor de expressão pROCK, a fim de verficar se o fenótipo das linhagens nocautes é revertido com o retorno do gene codificante para a TcRBP99. Para isso a sequencia correspondente a região codificadora do gene da TcRBP99 contendo *tag* de HA na sua porção C-terminal foi amplificada a partir do DNA genômico do clone CL Brener usando iniciador contendo a sequencia da *tag* de HA. O produto de PCR foi então clonado no vetor pROCK, dando origem ao plasmídeo pROCK_TcRBP99::HA (Figura 13A e 13B), o qual contem sequencia de beta-tubulina, do promotor de rRNA, além de sequencias necessárias para o correto processamento dos genes de puromicina e da TcRBP99. Após a linearização deste plasmídeo com a enzima de restrição *Not*I, parasitos nocautes para a TcRBP99 foram transfectados (Figura 13C) e selecionados com a droga puromicina, já que as drogas neomicina e higromicina já haviam sido utilizadas para selecionar a linhagem nocaute. Análise de PCR utilizando diferentes iniciadores mostrou que nos clones re-expressando a TcRBP99 as sequencias pKO_TcRBP99_Neo (Figura 13D) e pKO_TcRBP99_Higro (Figura 13E) estão integradas no genoma. Iniciadores que amplificam a região codificadora da TcRBP99 mostrou banda correspondente ao tamanho do gene dessa RBP (Figura 13F). Ensaio de PCR em tempo real revelou a presença de mRNA correspondente a TcRBP99 (Figura 13G).

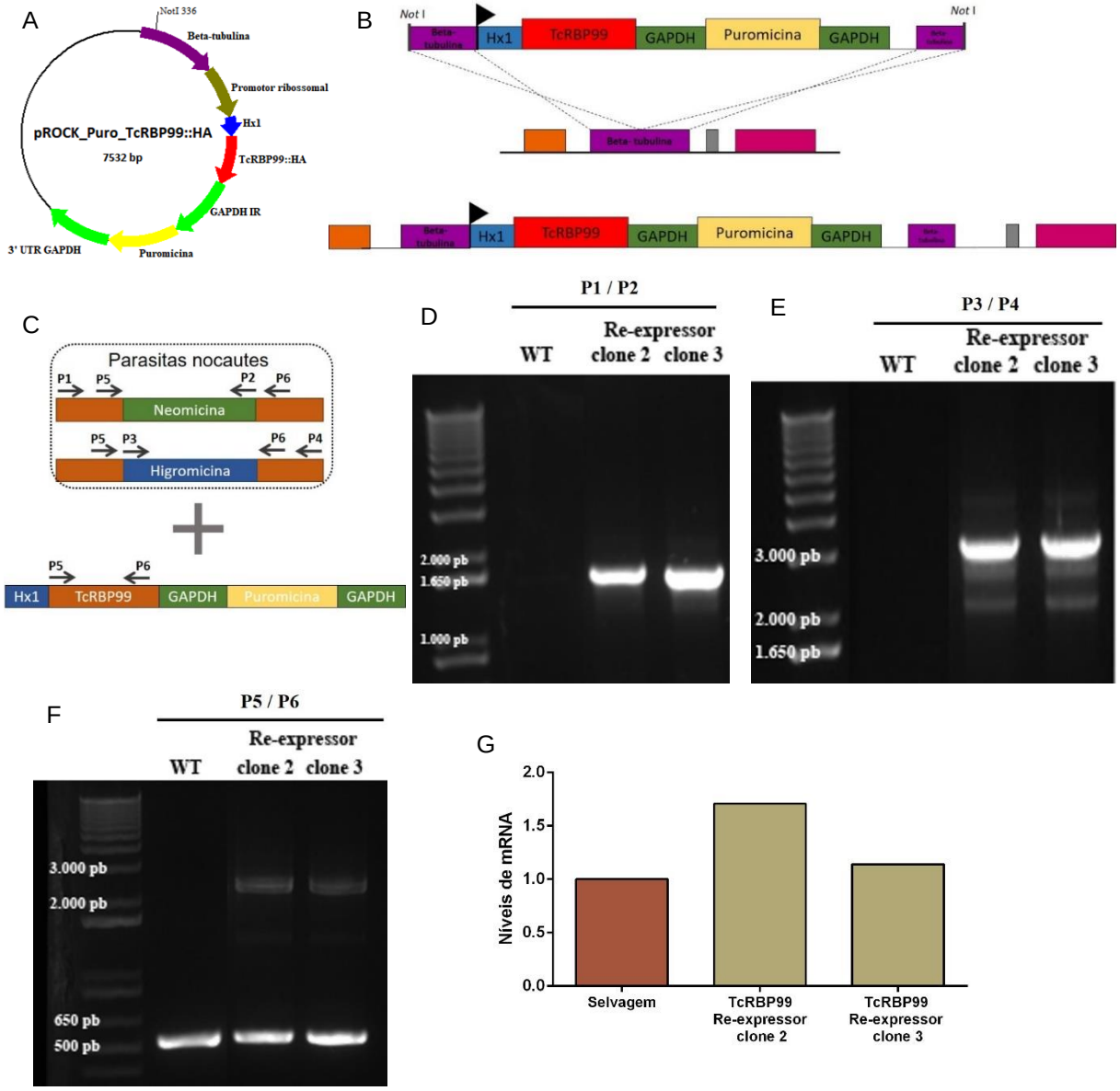


Figura 13: Construção do vetor para re-expressão do gene codificante para a TcRBP99 e obtenção de linhagens re-expressoras. Re-expressão do gene da TcRBP99 foi realizado utilizando o vetor de expressão em *T. cruzi* pROCK. (A) Região codificadora da TcRBP99 com *tag* de HA na porção C-terminal foi amplificada do genoma do clone CL Brener, previamente digerido com *Xba*I e *Xho*I para remoção da sequencia de GFP, dando origem ao plasmídeo pROCK_Puro_TcRBP99::HA. (B) Representação esquemática da integração do pROCK_Puro_TcRBP99::HA no *locus* de beta-tubulina após a linearização do plasmídeo com a enzima *Not*I. (C) Representação esquemática da transfecção de linhagens de parasitos nocautes com o plasmídeo pROCK_Puro_TcRBP99::HA. Setas indicam o local de anelamento dos iniciadores usados para confirmar a re-expressão do gene da TcRBP99 através de ensaios de PCR usando DNA genômico de parasitos selvagens e de dois clones re-expressando a TcRBP99 (D) O conjunto de iniciadores P1/P2 mostram a inserção da construção contendo neomicina no *locus* da TcRBP99. Fragmento esperado de 1.648 pares de bases. (E) Os iniciadores P3/P4 indicam a presença de higromicina no *locus* da TcRBP99. Fragmento esperado de 2.479 pares de bases. (E) Iniciadores para amplificar a região codificadora do gene da TcRBP99 (P5 e P6) mostram a presença da banda correspondente ao alelo da TcRBP99 que possui 522 pares de bases. (F) PCR em tempo real com iniciadores específicos para a TcRBP99.

4.6 Geração de linhagens super expressando a TcRBP99

Da mesma maneira como a re-expressão da TcRBP99 em parasitos nocautes, a super expressão dessa RBP foi realizada com o vetor de expressão em *T. cruzi* pROCK. A região codificadora da TcRBP99 contendo *tag* de HA na sua porção C-terminal foi amplificada a partir do DNA genômico do clone CL Brener e em seguida clonado no vetor pROCK, dando origem ao plasmídeo pROCK_Neo_TcRBP99::HA (Figura 14A). Diferentemente do plasmídeo pROCK_Puro_TcRBP99::HA, descrito no item anterior que contem resistência a puromicina, a seleção de população e clones de parasitos super expressando a TcRBP99 foi realizado com a droga G418 uma vez que o vetor usado continha sequencia do gene de resistência a neomicina. Com o intuito de confirmar a integração do plasmídeo no *locus* de beta-tubulina e consequente expressão da TcRBP99, western blot foi realizado utilizando extrato proteico total da população, de três clones e de parasitos selvagens. Como pode ser observado na figura 14B, além da população transfectada e selecionada com neomicina, apenas no clone 3 foi observada banda do tamanho esperado à TcRBP99 (aproximadamente 20 kDa). Ensaios de PCR em tempo real mostrou aumento de 1,5 vezes nos níveis de mRNA referentes a TcRBP99 no clone 3 (Figura 14C).

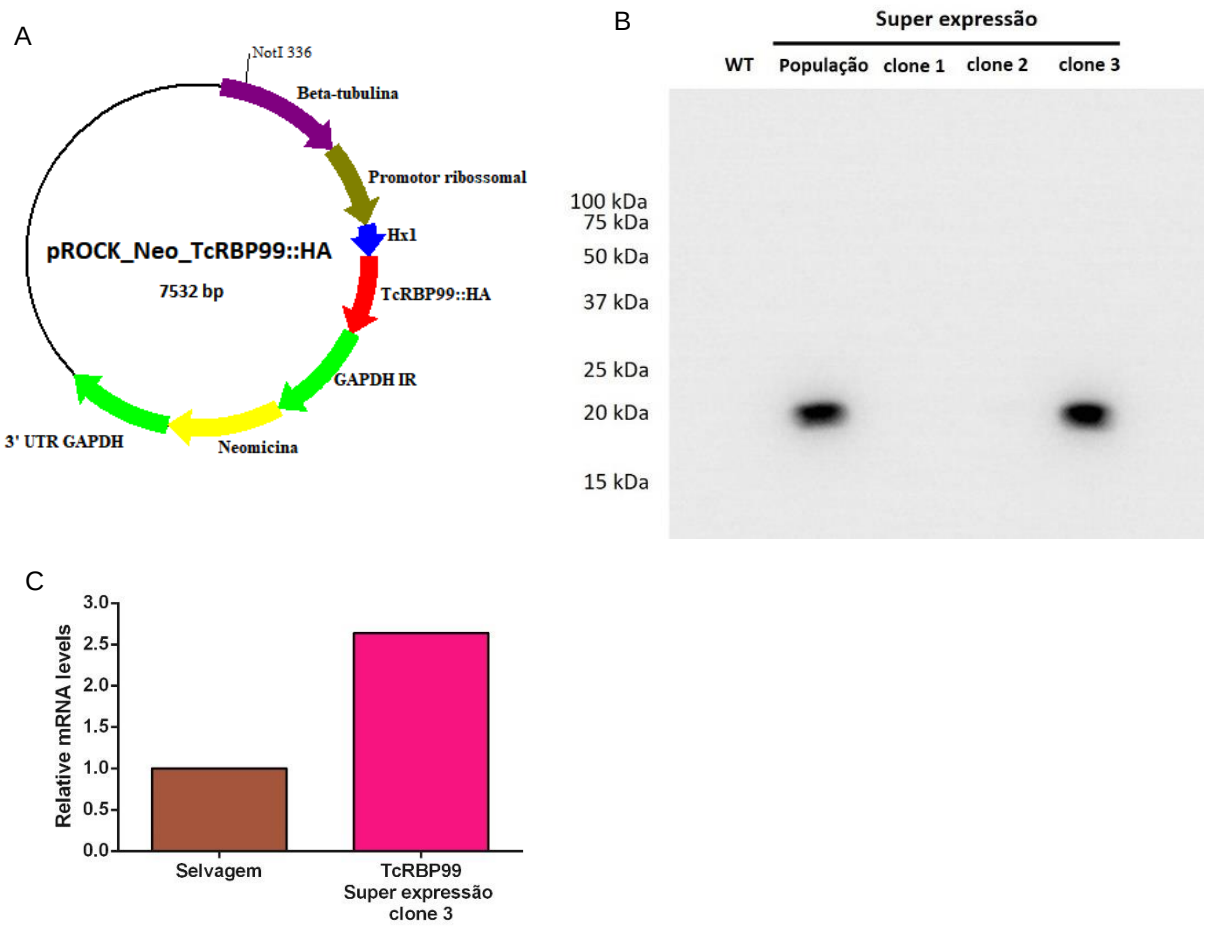


Figura 14: Construção do vetor para super expressar a TcRBP99 e obtenção de linhagens super expressoras. Super expressão do gene da TcRBP99 foi realizado utilizando o vetor de expresso em *T. cruzi* pROCK. (A) Região codificadora da TcRBP99 com *tag* de HA na porção C-terminal foi amplificada do genoma do clone CL Brener, previamente digerido com *Xba*I e *Xho*I para remoção da sequencia de GFP, dando origem ao plasmídeo pROCK_Neo_TcRBP99::HA. Parasitos CL Brener selvagens foram transfectados com esse plasmídeo, previamente linearizado com *Not*I, e população e clones foram selecionados com G418. (B) Wstern blot com extrato proteico total de $1x10^6$ parasitos selvagens, da população e de três clones super expressando a TcRBP99. Após a transferência dos extratos para membrana de nitrocelulose, a membrana foi incuba com anticorpo anti-HA na diluição de 1:5.000 e no dia segunda foi incubada com anticorpo anti-IgG de camundongo diluído 1:3.000. (C) PCR em tempo real com iniciadores específicos para a TcRBP99.

4.7 Caracterização da TcRBP99

Visando investigar o papel da proteína de ligação a RNA TcRBP99, a qual é altamente expressa em formas epimastigotas do *T. cruzi* comparado com as formas presentes no hospedeiro vertebrado, iniciamos o estudo gerando linhagens nocautes, como descrito nas sessões anteriores. Visto que essas linhagens nocautes foram viáveis, podemos concluir que, apesar de ter sua expressão fortemente aumentada em formas epimastigotas, essa RBP não é essencial para a sobrevivência do parasito nesse estágio do ciclo, pelo menos nas condições de cultivo de laboratório. Para a caracterização do fenótipo decorrente do nocaute desse gene, foram utilizados dois clones nocautes e dois clones re-rexpressores, ou seja, linhagens nocautes que foram re-transfectadas com o gene TcRBP99. Como controle foram utilizados parasitos CL Brener selvagens e parasitos transfectados com o vetor de expressão pROCK contendo apenas o gene de resistência, denominados pROCK vazio, gentilmente cedido pela Dra. Caroline Junqueira do Instituto René Rachou (Figura 15).

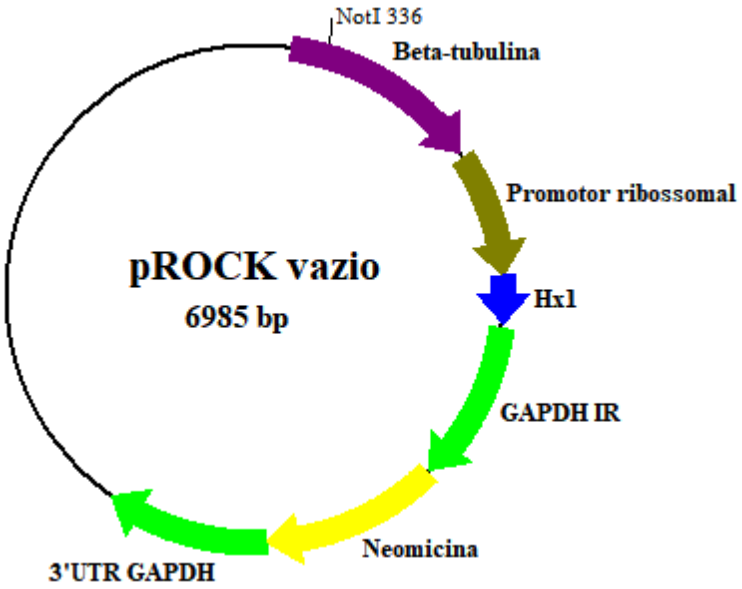


Figura 15: Plasmídeo pROCK vazio. Sequencia referente ao gene GFP foi removido do vetor de expressão em *T. cruzi*, pROCK, gerando o plasmídeo pROCK vazio, que foi linearizado com a enzima de restrição *Not*I e então usado para transfectar parasitos CL Brener selvagens. População desta linhagem foi selecionada com a droga G418

O crescimento das culturas de epimastigotas das linhagens obtidas foi acompanhado durante 9 dias, partindo de uma densidade inicial de $2x10^6$ parasitos/mL de cultura, a fim de verificar se a ausência da TcRBP99 interfere na capacidade de multiplicação dos parasitos. Ambos os clones de parasitos nocautes apresentaram significativa redução na taxa de crescimento 7 dias após o início da curva, quando estes clones são comparados aos parasitos selvagens (Figura 16A e 16B). As curvas de crescimento dos clones re-expressores foram similares as curvas de crescimento dos parasitos selvagens e da população transfectada com o vetor pROCK vazio ao longo de 7 dias, mas no nono dia, também apresentaram redução no crescimento (Figura 16C e 16D). Ao contrário do observado com os clones nocautes, a super expressão da TcRBP99 resultou em um aumento estatisticamente significativo no número de parasitos, em comparação aos selvagens, a partir do sétimo dia (Figura 16E e 16F). Não foram observadas diferenças entre o crescimento de parasitos CL Brener selvagens e da população transfectada com pROCK vazio (Figura 16C e 17E). Essses dados indicam que a TcRBP99 deve interagir e regular mRNAs importantes para o crescimento de epimastigotas do *T. cruzi*.

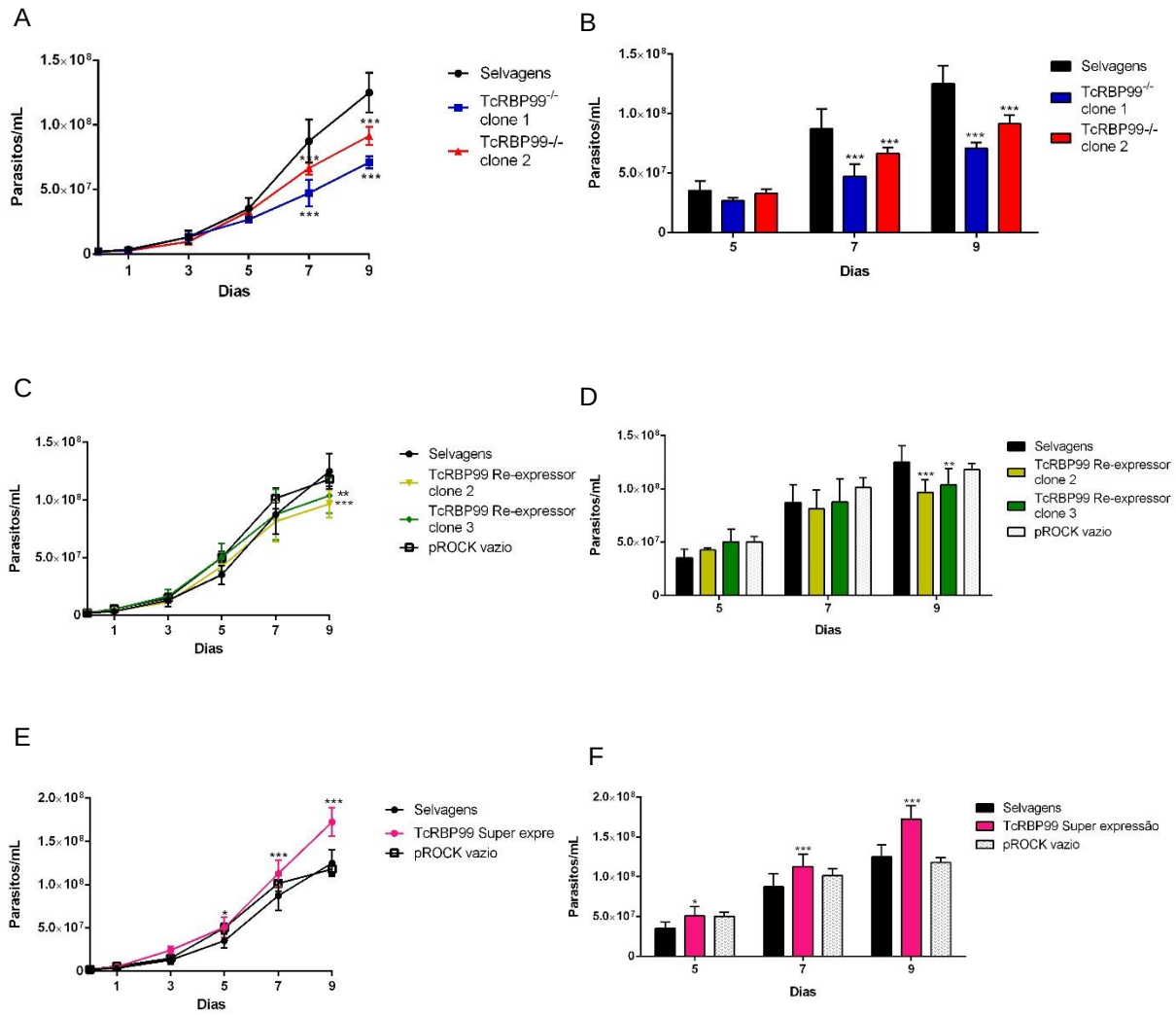


Figura 16: Curva de crescimento das diversas linhagens. Formas epimastigotas em fase logarítmica de crescimento foram acompanhadas por 9 dias. (A e B) Crescimento de parasitos nocautes (clone 1 em azul e clone 2 em vermelho) e CL Brener selvagens, representados em preto. (C e D) Crescimento de linhagens re-expressoras (clone 2 em amarelo e clone 3 em verde), parasitos selvagens representados por círculo fechado e pROCK vazio pelo quadrado aberto ou barra branca. (E e F) Crescimento do clone 3 super expressando a TcRBP99 (rosa), parasitos selvagens representados por círculo fechado e pROCK vazio pelo quadrado aberto ou barra branca. Dados representam média de quatro contagens. Os asteriscos indicam diferença significativa em análise two-way ANOVA, sendo *p<0,05, **p<0,01 e p***<0,001.

A capacidade das formas epimastigotas de se diferenciarem nas formas infectivas tripomastigotas metacíclicas, presentes no hospedeiro invertebrado, também foi usada para a caracterizar fenótipo das linhagens nocautes e re-expressoras. Para isso formas epimastigotas em fase logarítmica de crescimento foram transferidas para o meio RPMI, que induz a metaciclogênese por ser um meio pobre em nutrientes, e após 4 dias a porcentagem de tripomastigotas metacíclicas nas diferentes linhagens foi avaliada nas diferentes linhagens. Como pode ser observado na figura 17A, quatro dias após a indução da metaciclogênese um aumento na capacidade de diferenciação já era observado nos clones nocautes em comparação aos parasitos selvagens. Após 6 dias, aproximadamente 30% dos parasitos nocautes já haviam se diferenciado em tripomastigotas metacíclicas, enquanto que nos parasitos selvagens a porcentagem das formas infectivas foi de aproximadamente 20% após 10 dias de diferenciação. Clones re-rexpressores foram capazes de reverter essa alteração na capacidade de diferenciação em tripomastigotas metacíclicas apresentada pelos clones nocautes, comportando-se como os parasitos selvagens (Figura 17B). Por outro lado, o clone super expressando a TcRBP99 apresentou diminuição da capacidade de diferenciação em tripomastigotas metacíclicas após 6 dias, comparado ao clone CL Brener selvagem (Figura 17C). As análises com a linhagem transfectada com o vetor pROCK vazio não mostraram diferenças na taxa de metaciclogênese comparada à de parasitos selvagens e re-expressores (Figura 17B e 17C). Esses dados de metaciclogênese indicam que, nas formas epimastigotas, a TcRBP99 deve interagir com mRNAs envolvidos no processo de diferenciação de epimastigotas para tripomastigotas metacíclicas.

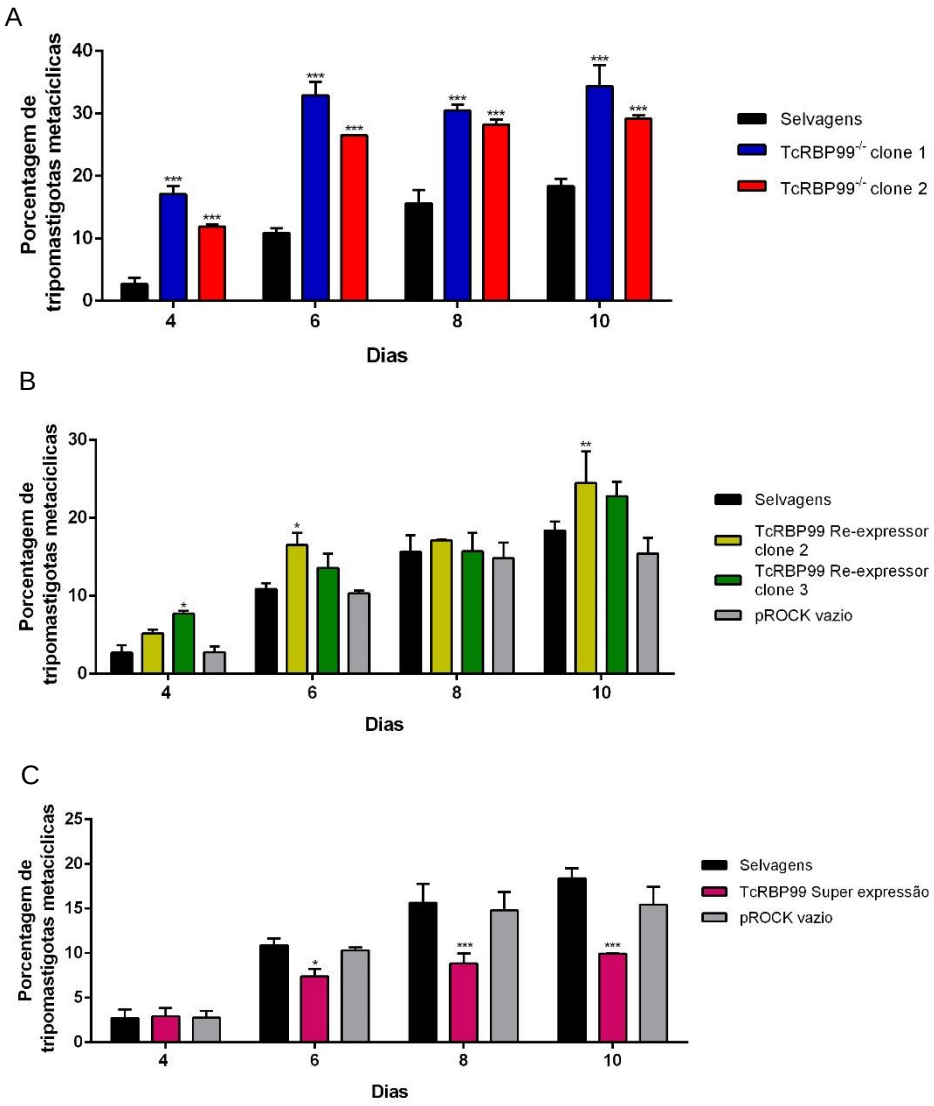


Figura 17: Metaciclogênese das diversas linhagens. Formas epimastigotas em fase logarítmica de crescimento foram retiradas do meio LIT e colocadas em meio RPMI. Após quatro dias as quantidades de tripomastigotas metacíclicas foram contadas a cada 2 dias durante 10 dias. (A) Porcentagem de tripomastigotas metacíclicas em parasitos nocautes (clone 1 e clone 2 em azul e vermelho respectivamente) e CL Brener selvagens, representado pela barra preta. (B) Porcentagem de tripomastigotas metacíclicas em linhagens re-expressoras (barras verdes representam o clone 2 e as amarelas o clone 3), em parasitos selvagens representados por barras pretas círculo e pROCK vazio pelas cinzas. (C) Porcentagem de tripomastigotas metacíclicas no clone super expressando a TcRBP99 (barras rosas), em parasitos selvagens representados por barras pretas círculo e pROCK vazio pelas cinzas. Dados representam média de duas contagens. Os asteriscos indicam diferença significativa em análise two-way ANOVA, sendo *p<0,05, **p<0,01 e p***<0,001.

4.8 Localização celular da TcRBP99

A localização celular da TcRBP99 foi determinada por meio da técnica de imunofluorescência. Para isso foram utilizadas formas epimastigotas de linhagens super expressando a TcRBP99 com *tag* de HA em sua porção C-terminal, que possibilitou a marcação da *tag* de HA com a fluorescência verde. Microscopia de fluorescência revelou que a proteína de ligação a RNA, TcRBP99, encontra-se dispersa por todo o citoplasma das formas epimastigotas (Figura 18).

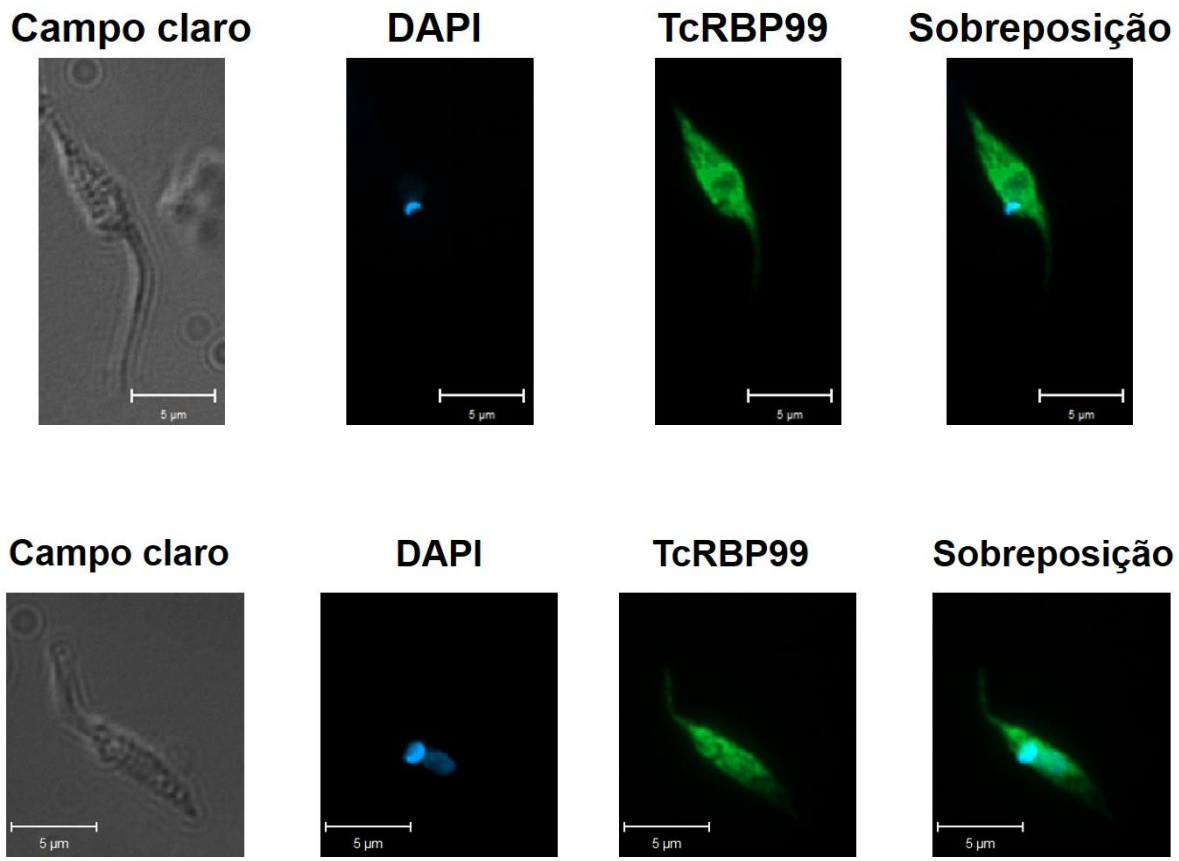


Figura 18: Localização celular da TcRBP99. Formas epimastigotas super expressando a proteína de ligação a RNA foram marcadas com anticorpo anti-HA e em seguida com anticorpo anti-rato conjugado com Alexa 488. Núcleo e kinetoplasto foram marcados com DAPI. Imagens foram capturadas no microscópio óptico de fluorescência Axio Imager Z2 – ApoTome 2-Zeiss.

4.9 RNA-Seq de linhagens nocautes para TcRBP99

Como descrito anteriormente, a ausência da TcRBP99 afetou o crescimento de epimastigotas, bem como sua capacidade de sua diferenciação em tripomastigotas metacíclicas, indicando que essa RBP poderia controlar os níveis de mRNAs envolvidos no crescimento e na diferenciação de epimastigotas. A fim de identificar quais mRNAs poderiam ser alvos da regulação pela TcRBP99, análises de expressão gênica global dos parasitos nocautes comparadas à de parasitos selvagens foram realizadas por meio do sequenciamento de bibliotecas de cDNA derivadas dos mRNAs obtidos das culturas de epimastigotas. Após a extração do RNA total de parasitos selvagens e dos dois clones nocautes, bibliotecas de cDNA foram construídas e sequenciadas na plataforma MiSeq. O processamento e análises dos dados dos transcriptomas foram realizados pela aluna de mestrado Wanessa Moreira Goes. Em resumo, foram geradas aproximadamente 3.500.000 *reads* para cada biblioteca que foram então submetidas a avaliação e "trimagem", sendo neste passo, descartadas em torno de 500 mil *reads*. Utilizando o programa TopHat2 aproximadamente 70% a 75% das *reads* foram mapeadas nos dois haplótipos do clone CL Brener (Tabela 5).

Tabela 5: Resumo da quantidade de *reads* geradas mapeadas nas sete bibliotecas.

Biblioteca	**Total de *reads***	**Total de *reads* após trimagem**	***Reads* mapeadas (%)**
Selvagem	3.294.361	2.709.211	72,5%
Selvagem	3.609.438	2.862.761	74,9%
Selvagem	3.469.573	2.626.697	75,0%
Nocaute	3.513.887	2.741.084	70,2%
Nocaute	3.895.258	3.088.360	72,9%
Nocaute	4.143.756	3.274.477	73,1%
Nocaute	3.521.679	2.820.681	74,4 %

Na tabela 6 estão descritos os genes diferencialmente expressos, considerando valores de log2 *fold change* maiores ou menores que 1 e *p-value* ajustado menor que 0,05, entre os parasitos selvagens e os nocautes para a TcRBP99. Nota-se que foram identificados 12 genes com expressão diminuída nos parasitos nocautes em relação aos selvagens. Genes com expressão aumentada nos parasitos nocautes comparados aos selvagens não foram identificados com base nas análises estatísticas utilizadas, o que sugere que a TcRBP99 poderia ser uma proteína que se liga a mRNAs estabilizando-os. Dos 12 genes com expressão aumentada nos parasitos selvagens comparados aos clones nocautes, encontramos três transcritos pertencentes a família dos retrotransposons hot spot (RHS), um gene codificador de um transportador de aminoácidos, um gene codificador de uma proteína associada a diferenciação (PAD) e seis genes codificadores de proteínas hipotéticas. Como foram encontradas *reads* que mapearam no gene codificador da proteína de ligação a RNA TcRBP99, apesar de haver um número menor de *reads* nos nocautes em comparação aos parasitos selvagens, podemos supor que isso ocorreu devido ao fato da região 5'UTR correspondente ao gene da TcRBP99 bem como o início da região codificadora do gene terem permanecidos no genoma dos clones nocautes. Assim, essa região foi transcrita e as *reads* geradas alinharam nessa região. Como esperado, não foram mapeadas *reads* na região do gene da TcRBP99 interrompida pela sequencia Hx1_Neo_GAPDH.

Tabela 6: Genes com expressão diminuída na linhagen nocaute comparado aos parasitos selvagens.

ID	Anotação no genoma	log2 *fold change*	p-value ajustado
TcCLB.509559.20	retrotransposon hot spot protein (RHS, pseudogene)	-2,86223	1,06E-35
TcCLB.505943.110	retrotransposon hot spot protein (RHS, pseudogene)	-2,74450	1,52E-32
TcCLB.506739.99	zinc finger (CCCH type) protein	-1,6018	6,66E-35
TcCLB.503733.80	hypothetical protein, conserved	-1,50603	1,42E-19
TcCLB.506153.10	amino acid transporter	-1,26987	3,04E-08
TcCLB.506551.10	protein associated with differentiation 8	-1,14790	1,01E-14
TcCLB.504173.60	hypothetical protein	-1,14374	1,04E-07
TcCLB.504163.40	hypothetical protein	-1,12380	3,65E-05
TcCLB.506825.40	hypothetical protein	-1,12250	6,19E-05
TcCLB.507659.10	hypothetical protein	-1,05995	4,13E-06
TcCLB.511185.150	hypothetical protein (pseudogene)	-1,01498	0,000622783
TcCLB.511843.80	retrotransposon hot spot (RHS) protein	-1,01447	2,80E-05

A expressão de genes codificantes para transportadores de aminoácidos é importante nas formas epimastigotas uma vez que essas formas têm como principal fonte de energia os aminoácidos. No genoma de CL Brener encontramos 17 genes codificantes para os transportadores de aminoácidos e praticamente todos eles, de acordo com dados de transcriptoma de CL Brener, apresentam expressão aumentada nas formas epimastigotas, como pode ser observado na figura 19A. Diminuição nos níveis de mRNA do gene TcCLB.506153.10, codificante para um membro da família de transportadores de aminoácidos, nos parasitos nocautes pode ser observada no mapeamento das *reads* (Figura 19B). Redução de 70% dos níveis de mRNA nos nocautes foi observada por PCR em tempo real (Figura 19C). Linhagens re-expressando o gene da TcRBP99 apresentaram níveis de mRNA correspondentes ao transportador de aminoácido similares aos níveis presente nos parasitos selvagens. Esses dados indicam que a TcRBP99 regula a expressão do gene codificador para o

transportador de aminoácidos, já que parasitas nocautes para essa RBP apresentam redução nos níveis de expressão do gene codificante para o transportador de aminoácidos em epimastigotas.

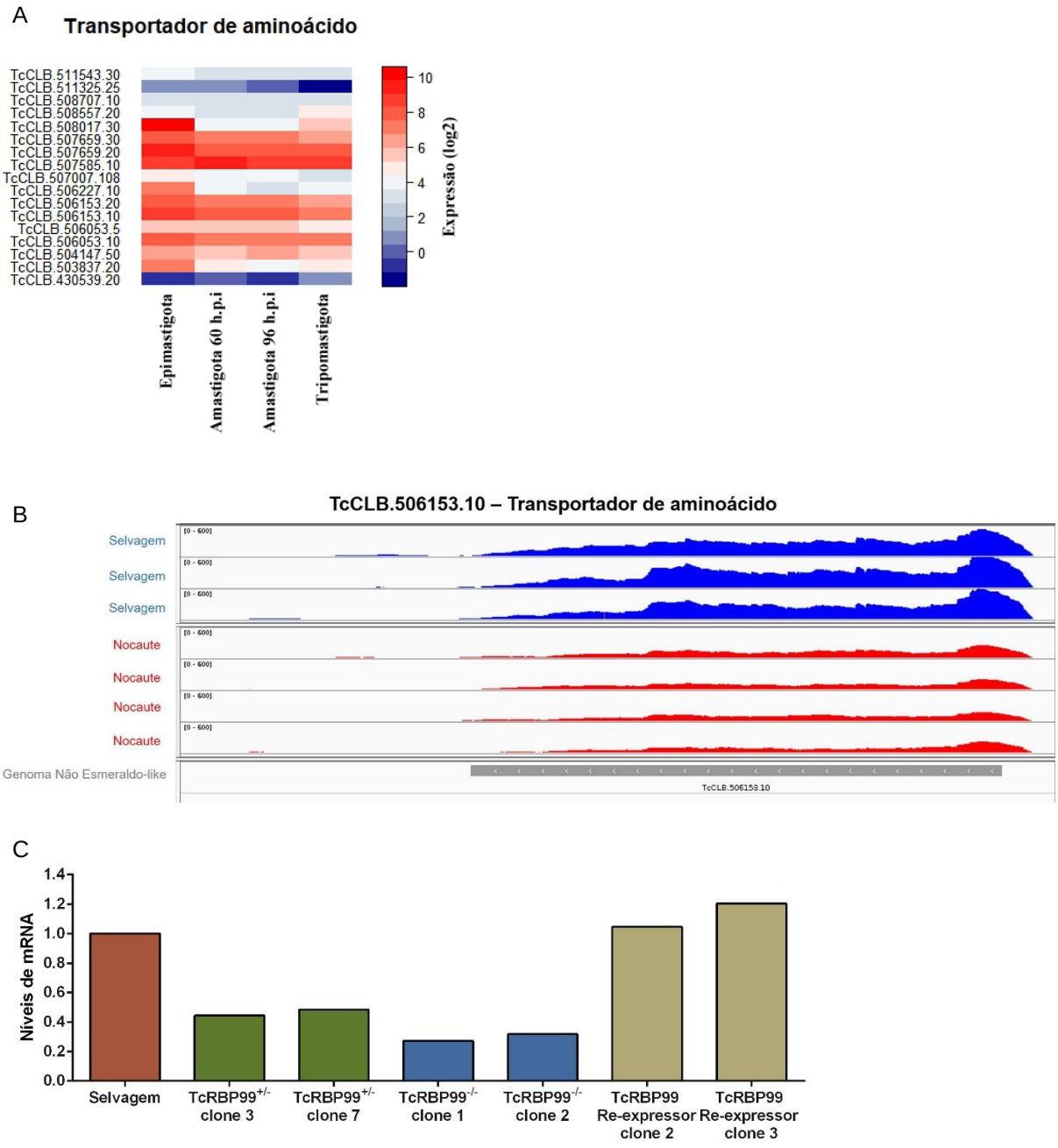


Figura 19: Gene codificantes para transportador de aminoácidos. Transcriptoma de linhagens selvagens e nocautes revelaram a presença de um gene codificante para um transportador de aminoácido entre os genes diferencialmente expressos. (A) *Heatmap* com a expressão dos genes codificantes para transportadores de aminoácidos nas formas epimastigota, amastigota coletadas 60 e 96 horas após a infecção e tripomastigotas liberadas em cultura. (B) Mapeamentos das *reads* nas bibliotecas correspondes as linhagens selvagens e nocautes no transcrito TcCLB.506153.10. (C) PCR em tempo real com iniciadores específicos para o transportador de aminoácido.

O gene codificante para as proteínas associadas a diferenciação (PAD) apresentam 8 cópias no genoma de CL Brener. A figura 20A mostra o mapeamento das *reads* na região correspondente ao gene TcCLB.506551.10, que codifica uma proteína associada a diferenciação, onde é possível observar um número menor de *reads* nos mapeamentos correspondentes as linhagens nocautes. De acordo com os dados de RNA-seq, todas as 13 sequencias de genes codificantes para as PADs também possuem expressão aumentada em epimastigotas de linhagens selvagens em comparação às formas encontradas no hospedeiro vertebrado (Figura 20B). Análises por PCR em tempo real usando iniciadores capazes de se ligarem à uma região conservada entre membros da PAD, revelou diminuição da expressão global dos mRNAs codificantes para as proteínas associadas a diferenciação nas linhagens nocautes. Nas linhagens hemi-nocautes e re-rexpressoras não foi observado diferença entre os níveis de mRNAs comparado aos parasitos selvagens (Figura 20C). Assim como para o gene codificante para o transportador de aminoácidos, esses dados indicam que a TcRBP99 regula a expressão do gene codificador para a proteína associada a diferenciação, já que parasitas nocautes para essa RBP apresentam redução parcial nos níveis de expressão do gene codificante para a PAD em epimastigotas.

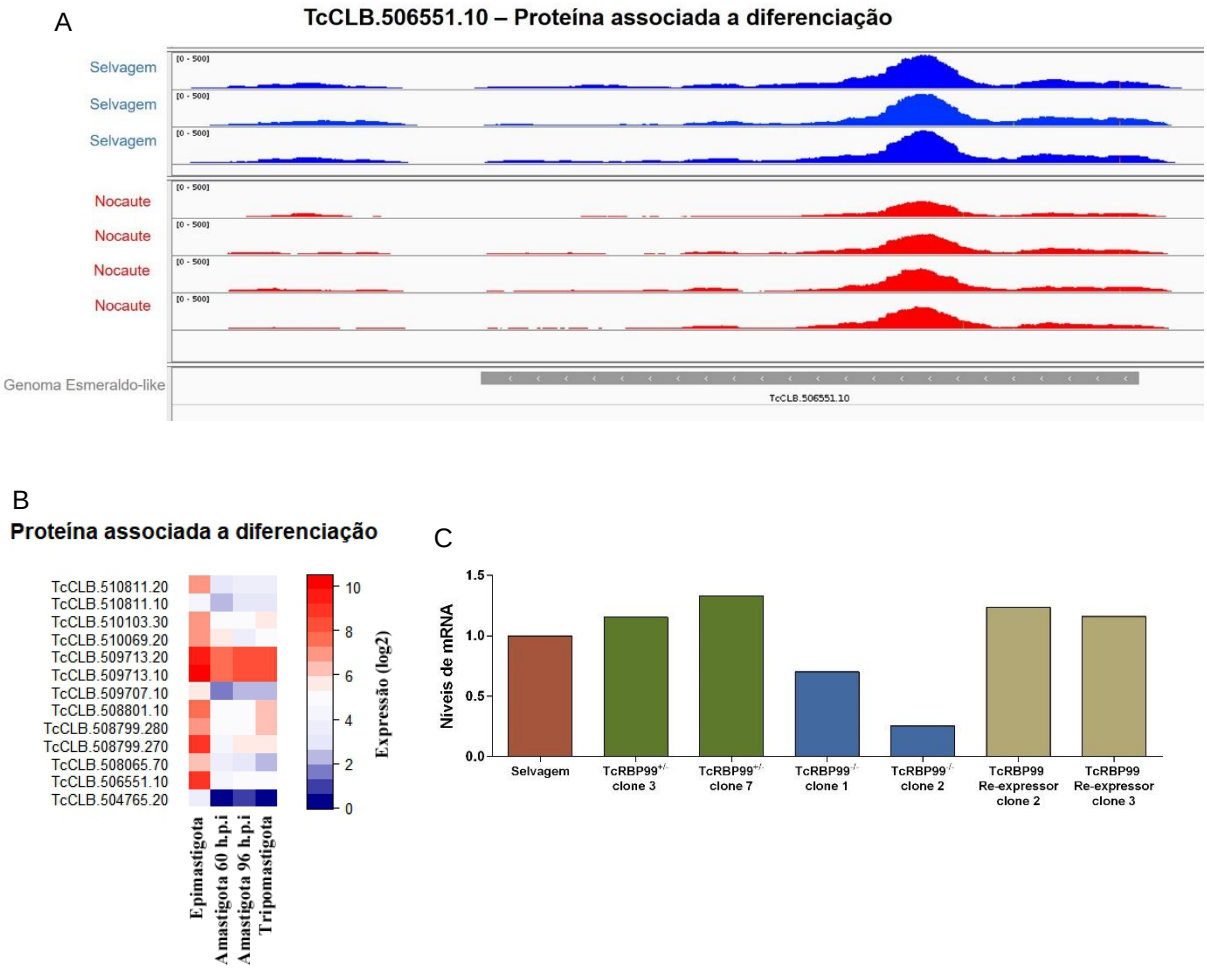


Figura 20: Gene codificante para proteína associada a diferenciação. Transcriptoma de linhagens selvagens e nocautes revelaram a presença de um gene codificante para uma proteína associada a diferenciação entre os genes diferencialmente expressos. (A) Mapeamentos das *reads* nas bibliotecas correspondes as linhagens selvagens e nocautes no transcrito TcCLB.506551.10. (B) *Heatmap* com a expressão dos genes codificantes para proteínas associadas a diferenciação nas formas epimastigota, amastigota coletadas 60 e 96 horas após a infecção e tripomastigotas liberadas em cultura. (E) PCR em tempo real com iniciadores para região conservada entre os membros das PADs.

O fato de terem sido identificados membros da família dos retrotransposons hot spot (RHS), uma das famílias multigênicas identificadas durante a montagem do genoma do clone CL Brener, entre os genes com expressão diminuída nos clones nocautes sugere que esses genes também possam ser alvos da TcRBP99. Como mostra a Figura 19, uma clara redução nos níveis de expressão dos membros das RHS nos nocautes pode ser observada no mapeamento das *reads* (Figuras 21A a 21C). Análise de *heatmap* com dados do RNA-Seq das formas epimastigotas, amastigotas coletadas 60 e 96 horas após a infecção e tripomastigotas, mostrou que a maior parte dos membros da família multigênicas dos RHS tem expressão aumentada em epimastigotas (Figura 21D). A fim de confirmar se houve uma redução geral da expressão da família RHS nos parasitos nocautes, foi realizada PCR em tempo real com iniciadores que anelam em uma região comum aos membros da família RHS. Resultados mostraram que não houve alteração na expressão geral dos membros da família RHS, nas linhagens hemi-nocautes, nocautes e re-expressoras (Figura 21E). Considerando que as famílias multigênicas apresentam regiões conservadas, a ausência de diferença de expressão observada na RT-PCR, entre linhagens nocautes e selvagens, pode ser ter ocorrido devido a presença de mRNAs codificantes para outros membros das RHS nas linhagens nocautes.

Sequencias das proteínas hipotéticas identificadas na análise de expressão diferencial foram submetidas a pesquisas para identificação de domínios nos bancos de dados Pfam e InterPro. O domínio RING variante, que é um motivo *zinc finger-like* (C_4HC_3), foi identificado na proteína hipotética identificada como TcCLB.504173.60, indicando que essa proteína possa se ligar a mRNA. A proteína hipotética TcCLB.507659.10 possui o domínio SPFH, também conhecido como Band 7, que está associado com proteínas que se ligam à membrana. Não foram encontradas

sequencias conservadas entre as proteinas hipotéticas diferencialmente expressas nos parasitos nocautes comparado aos selvagens.

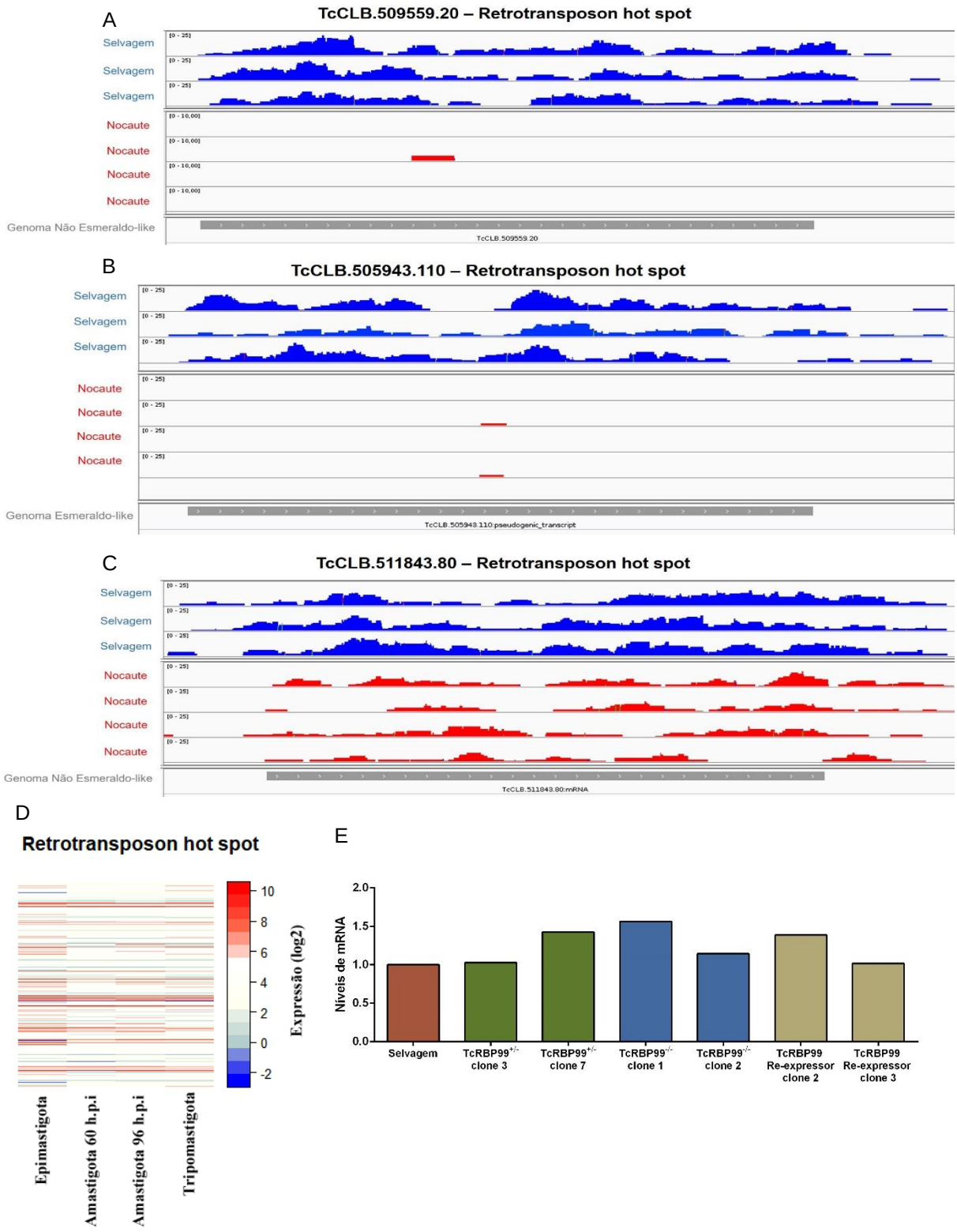


Figura 21: Genes codificantes para retrotransposons hot spot. Transcriptoma de linhagens selvagens e nocautes revelaram a presença de três membros da família das RHS entre os genes diferencialmente expressos. Mapeamentos das *reads* nas bibliotecas correspondes as linhagens selvagens e nocautes nos transcritos (A) TcCLB.509559.20, (B) TcCLB.509543.110 e (C) TcCLB.511843.80. (D) *Heatmap* com a expressão dos membros da família RHS nas formas epimastigota, amastigota coletadas 60 e 96 horas após a infecção e tripomastigotas com base nos dados de RNA-seq. (E) PCR em tempo real com iniciadores para região comum entre os membros das RHS.

4.10 Interação entre RNA mensageiros de genes diferencialmente expressos e a proteína de ligação a RNA TcRBP99

Sequenciamento de mRNAs de linhagens selvagens e nocautes permitiu identificar os genes que sofrem alteração de expressão quando a TcRBP99 é removida do genoma do clone CL Brener. A fim de verificar se a TcRBP99 interage com o mRNAs dos genes diferencialmente expressos entre essas duas linhagens, parasitos super expressando a TcRBP99 com *tag* de *HA*, foram utilizados em ensaios de imunoprecipitação, com anticorpos anti-HA, seguidos de extração de RNA total correspondente aos genes presentes no sobrenadante (não ligados a TcRBP99) e aos que foram eluidos com a TcRBP99, isto é que permaneceram ligados à TcRBP99 (Figura 22A). Ensaios de western blot nos quais a membrana contendo extrato proteico correspondente às frações total, sobrenadante e eluato foi incubada com anticorpo anti-HA mostram a presença da TcRBP99::HA, de aproximadamente 20kDA, nas frações total e eluato (Figura 22B). Bandas de 25 kDa e 50 kDa correspondente as cadeias leve e pesada, respectivamente, do anticorpo anti-HA ligado ao anticorpo secundário também podem ser visualizadas na membrana do western blot (Figura 22B).

Inicialmente verificamos se ocorreu a interação entre a TcRBP99 com o mRNA codificante para a proteína associada a diferenciação (gene TcCLB.506551.10). Para isso cDNA foi sintetizado a partir de RNA total extraído das três frações, e em seguida a região codificadora desta PAD foi amplificada por PCR. Como pode ser observado na figura 22C, não é observada banda correspondente à região codificadora da proteína associada a diferenciação (1.845 pares de base) na fração do eluato dos parasitos selvagens, utilizaados como controle da imunoprecipitação. Uma banda do tamanho esperado, de 1.845 pares de base, é observada nas frações total e eluato

de parasitos super expressando a TcRBP99::HA, indicando que a proteína TcRBP99 interage com o mRNA codificante para a proteína associada a diferenciação (Figura 22C). cDNAs sintetizados na ausência da enzima transcriptase reversa foram utilizados como controle negativo da reação de PCR, uma vez que sem essa enzima não há síntese do DNA de dupla fita, o que impossibilita a amplificação pela técnica de PCR. Como esperado, não foi observada banda na fração correspondente ao eluato dos parasitos super expressando a TcRBP99::HA (Figura 22D). Com o objetivo de verificar a especificidade de interação da TcRBP99 com o gene codificante para a PAD uma outra proteína de *T. cruzi* contendo *tag* de HA foi usada em ensaio de imunoprecipitação e a PAD foi amplificada a partir de cDNAs sintetizados das diversas frações. Nas figuras 22E e 22F observa-se a ausência da banda correspondente a região codificadora do gene codificante para a PAD na fração do eluato dos parasitos expressando a *tag* de HA, tanto na presença quanto na ausência da enzima transcriptase reversa. Com base nesses dados é possível concluir que a proteína de ligação a RNA, TcRBP99, interage com o mRNA codificante para a proteína associada a diferenciação (TcCLB.506551.10).

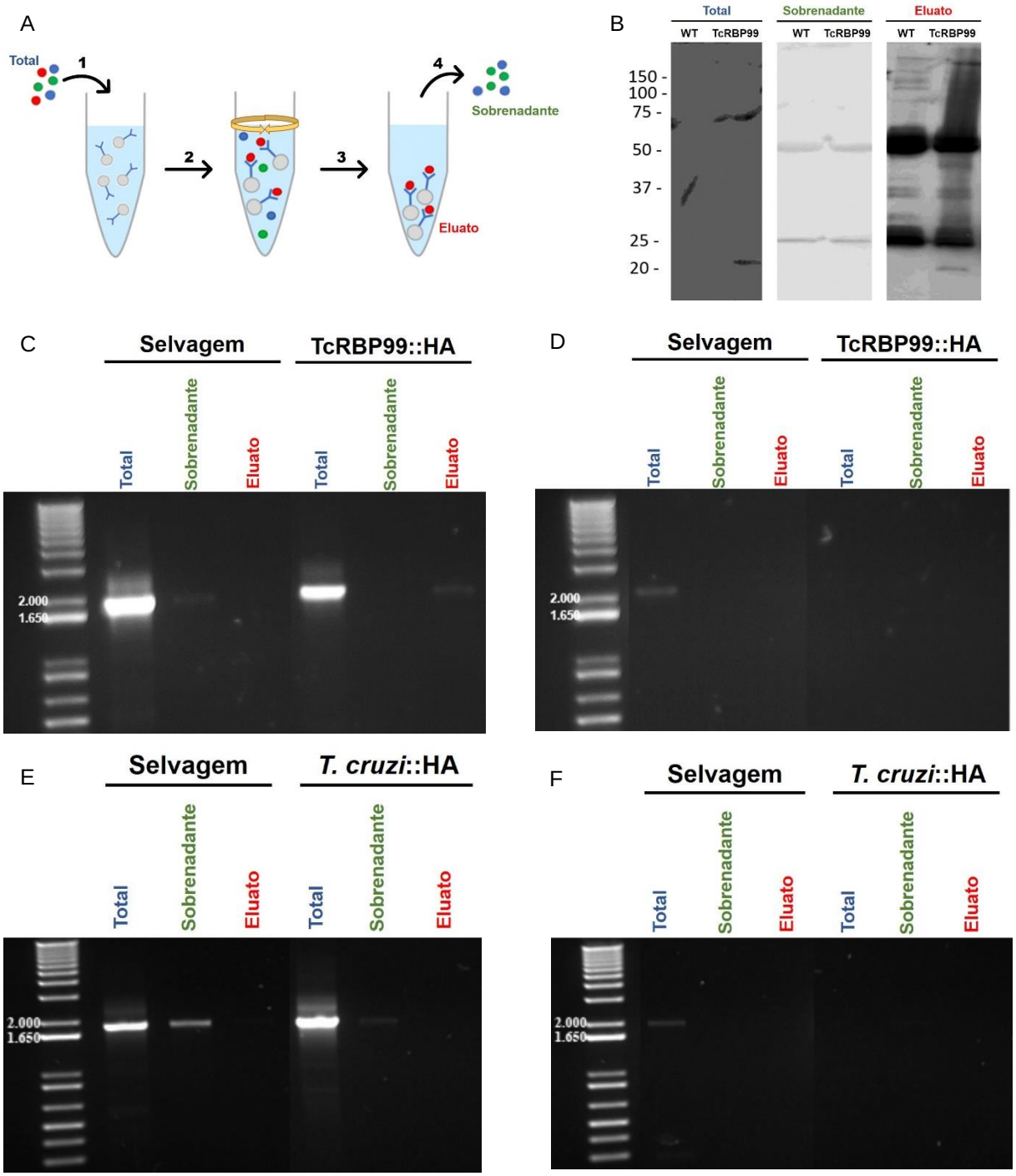


Figura 22: Interação da TcRBP99 e o mRNA codificante para uma proteína associada a diferenciação. Interação entre a TcRBP99 e o mRNA codificante para a proteína associada a diferenciação foi realizada usando parasitos super expressando a TcRBP99::HA. (A) Representação esquemática do processo de imunoprecipitação: 1- Parasitos selvagens e super expressando a TcRBP99::HA foram lisados e o sobrenadante foi adicionado a *beads* ligado a anticorpo anti-HA; 2- As *beads* ficaram incubando com o lisado por aproximadamente 4 horas para que a TcRBP99::HA se ligue ao anticorpo; 3- As *beads* foram centrifugadas para concentrar a TcRBP99::HA ; 4- O sobrenadante foi transferindo para um novo tudo. RNA total foi extraído das frações total (azul), sobrenadante (verde) e eluato (vermelho). (B) Western blot com extrato proteico das frações total, sobrenadante e eluato foi revelado com anticorpo anti-HA (1:5.000). Amplificação da região codificadora da PAD nas diversas frações da imunoprecipitação realizada com parasitos (C e D) super expressando a TcRBP99::HA e (E e F) expressando outra proteína de *T. cruzi* contendo *tag* de HA (*T. cruzi*::HA). cDNA foi sintetizado (C e E) na presença e (D e F) na ausência da enzima transcriptase reversa

4.11 Identificação dos motivos presentes na região 3'UTR de genes possíveis alvos de regulação da TcRBP99

Considerando que, de maneira geral, os motivos regulatórios reconhecidos pelas RBPs localizam-se nas regiões 3' UTR dos mRNAs alvos, sequencias correspondentes as regiões 3'UTR dos genes diferencialmente expressos entre os parasitos selvagens de CL Brener e linhagens nocautes para a TcRBP99 foram analisadas após a determinação dessa sequencias com base no mapeamento das *reads* do transcriptoma. Para isso, as regiões 3'UTR, que tem início na base seguinte ao códon de parada e termina nas últimas *reads* mapeadas antes do próximo gene, dos genes diferencialmente expressos foram extraídas do mapeamento e a sequencia de cDNA foi convertida em mRNA. Os tamanhos das regiões 3'UTR dos genes diferencialmente expressos foram estimados através do mapeamento e estão descritos na tabela 7. É possível observar nessa tabela que genes codificantes para proteínas hipotéticas possuem a menor e a maior região 3'UTR com 125 e 1.859 nucleotídeos respectivamente.

Tabela 7: Estimativa do tamanho da região 3'UTR dos genes com expressão diferencial entre parasitos selvagens e nocautes para a TcRBP99.

Identificador	Gene	Tamanho da região 3'UTR
TcCLB.505943.110	Retrotransposon hot spot	534 nucleotídeos
TcCLB.506551.10	Proteína associada a diferenciação	889 nucleotídeos
TcCLB.506825.40	Proteína hipotética	1.859 nucleotídeos
TcCLB.509559.20	Retrotransposon hot spot	171 nucleotídeos
TcCLB.503733.80	Proteína hipotética	1.031 nucleotídeos
TcCLB.506153.10	Transportador de aminoácidos	1.115 nucleotídeos
TcCLB.504173.60	Proteína hipotética	125 nucleotídeos
TcCLB.504173.40	Proteína hipotética	240 nucleotídeos
TcCLB.507659.10	Proteína hipotética	278 nucleotídeos
TcCLB.511843.80	Retrotransposon hot spot	548 nucleotídeos

A presença de motivos conservados nas regiões 3'UTR poderia indicar o local em que ocorre a interação entre a TcRBP99 e os mRNAs. Pesquisa por região conservada em todas as regiões 3'UTR dos mRNAs não mostrou presença de sequencias ou bases conservadas quando sequencias de todos os genes diferencialmente expressos foram analisadas em conjunto. Devido a isso, as sequências correspondentes aos mRNAs codificantes para os genes com expressão diferencial foram separados em três grupos e então submetidos a pesquisa de motivos. O primeiro grupo, que é formando pelos três genes codificantes para membros da família RHS, mostra através dos blocos coloridos que a região 3'UTR é conservada entre esses membros da RHS (Figura 23A) porém não se observa predominância de nucleotídeos (Figura 23B). No segundo grupo, composto pelos transcritos codificantes para proteínas hipotéticas, é observado presença de regiões conservadas que, no entanto, não estão situados na mesma sequência nas regiões correspondentes a 3'UTRs das proteínas hipotéticas analisadas (Figura 23C). Apesar

disso, a análise mostra uma alta frequência do nucleotídeo adenosina nas sequencias analisadas (Figura 23D). Por fim agrupando os genes codificantes para a proteína associada a diferenciação e para o transportador de aminoácidos observa-se que assim como para o grupo dos RHS, há conservação entre as regiões 3'UTR e que os motivos conservados se encontram na mesma posição nas duas sequencias analisadas, como pode ser observado pela disposição dos blocos coloridos na figura 23E. O motivo conservado entre essas duas sequencias apresenta alta frequência do nucleotídeo citosina (Figura 23F). Esses dados indicam que a TcRBP99 deve interagir com as bases adenina e citocina presentes na região 3'UTR dos mRNAs. O conjunto de resultados dos experimentos que caracterizam o papel da TcRBP99 foram reunidos em um manuscrito que sera submetido a publicação à revista Journal of Biological Chemistry (Apêndice IV).

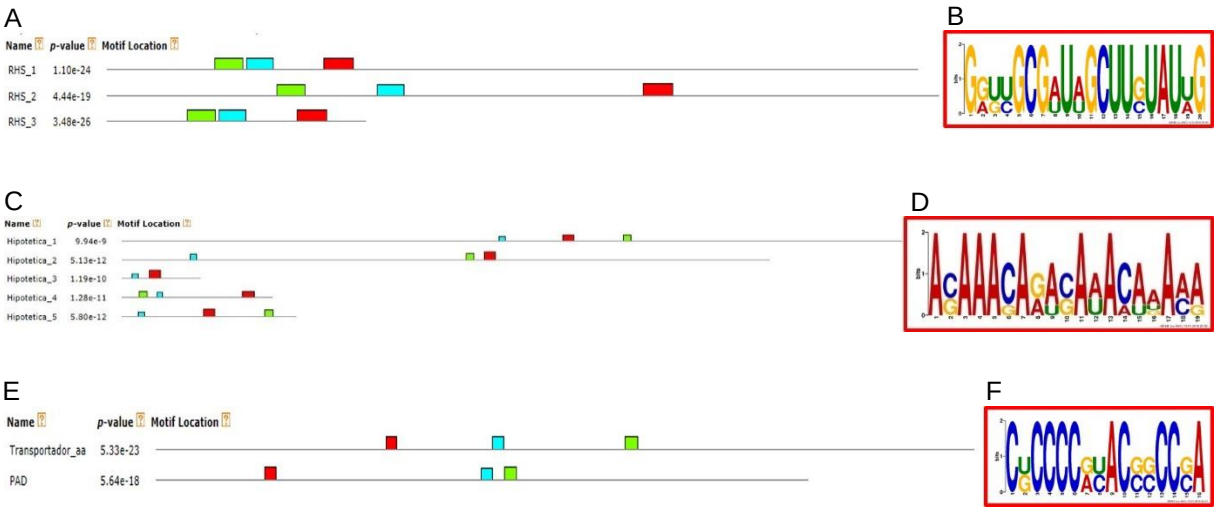


Figura 23: Motivos conservados presente na região 3'UTR dos genes com expressão diferencial entre parasitos selvagens e linhagens nocautes para TcRBP99. Região 3'UTR que é compreendida pela sequencia de bases que começa no primeiro nucleotídeo depois do códon de parada e termina nas ultimas *reads* mapeadas antes do próximo gene, foi identificada depois do mapeamento das *reads*. Sequencias correspondentes a região 3'UTR dos genes com expressão diferencial foram extraídas e após a conversão do cDNA em mRNA foram submetidas a pesquisa de motivos em grupos. (A) Motivos conservados entre os genes diferencialmente expressos codificantes para as RHS e (B) sequencia conservada entre os 3 membros da família RHS. (C) Motivos conservados entre os transcritos codificantes para as proteínas hipotéticas e (D) sequencia conservada entre as regiões 3'UTR das proteínas hipotéticas. (E) Motivos conservados entre os genes codificantes para a proteína associada a diferenciação e o transportador de aminoácido e (F) sequencia conservada presente na 3'UTR desses mRNAs. Os blocos coloridos indicam motivos conservados entre os mRNAs analisados.

5. RESULTADOS - PARTE II

5.1 Proteínas de ligação a RNA diferencialmente expressas entre cepas de *T. cruzi* apresentando diferentes graus de virulência

Além da pesquisar pelas proteínas de ligação a RNA diferencialmente expressas entre as formas que habitam os diferentes hospedeiros do *T. cruzi*, procuramos também por proteínas de ligação a RNA que poderiam explicar a diferença de virulência encontrada entre os clones CL Brener e CL-14 do *T. cruzi*. Para isso focamos na diferenciação intracelular que ocorre no hospedeiro vertebrado, uma vez que os fatores associados a virulência não devem estar envolvidos com as formas que habitam o hospedeiro invertebrado. Sendo assim, analisamos o transcriptoma de formas amastigotas coletadas 60 e 96 horas após a infecção e tripomastigotas liberadas em cultura provenientes dos clones CL Brener e CL-14 (Figura 24A). O gene codificante para a RBP identificada por TcCLB.511511.6 aparece entre as RBPs diferencialmente expressas em tripomastigotas, tanto no clone CL Brener quanto no CL-14, quando comparações entre as formas replicativa e as forma sanguínea foram realizadas (Figura 24B). Nesta mesma comparação, foram ainda identificados cinco transcritos (TcCLB.511727.190, TcCLB.469785.40, TcCLB.509551.60, TcCLB.504005.6, TcCLB.506563.10) com expressão aumentada em amastigotas nos dois clones (Figuras 24C a 24G). Buscando por mRNAs diferencialmente expressos entre os clones estudados, identificamos o gene TcCLB.507611.300 (Figura 24H) que apresenta expressão aumentada apenas nas formas tripomastigotas do clone avirulento CL-14. Análise da sequencia das proteínas de ligação a RNA diferencialmente expressas durante o ciclo intracelular nos clones CL Brener e CL-14 mostrou que essas RBPs diferencialmente expressas possuem entre 124 e 535 aminoácidos (Figura 24I). A RBP que apresentou expressão aumentada em

tripomastigota de ambos os clones possui o domino *zinc finger*, enquanto que nas proteínas de ligação a RNA que apresaentaram expressão aumentada em amastigotas, independente do tempo de infecção, foram identificados os motivos RRM, *zinc finger* e pumilio, como pode ser observado na figura 24I. O gene TcCLB.507611.300, codificante para a RBP que apresentoa expressão aumentada em tripomastigotas somente do clone CL-14, denominado TcRBP300 e que possui um domínio RRM, foi escolhido para estudos buscando caracterizar o seu papel como um fator possivelmente envolvido com as diferenças de virulência observadas entre as os clones CL Brener e CL-14 do *T. cruzi*.

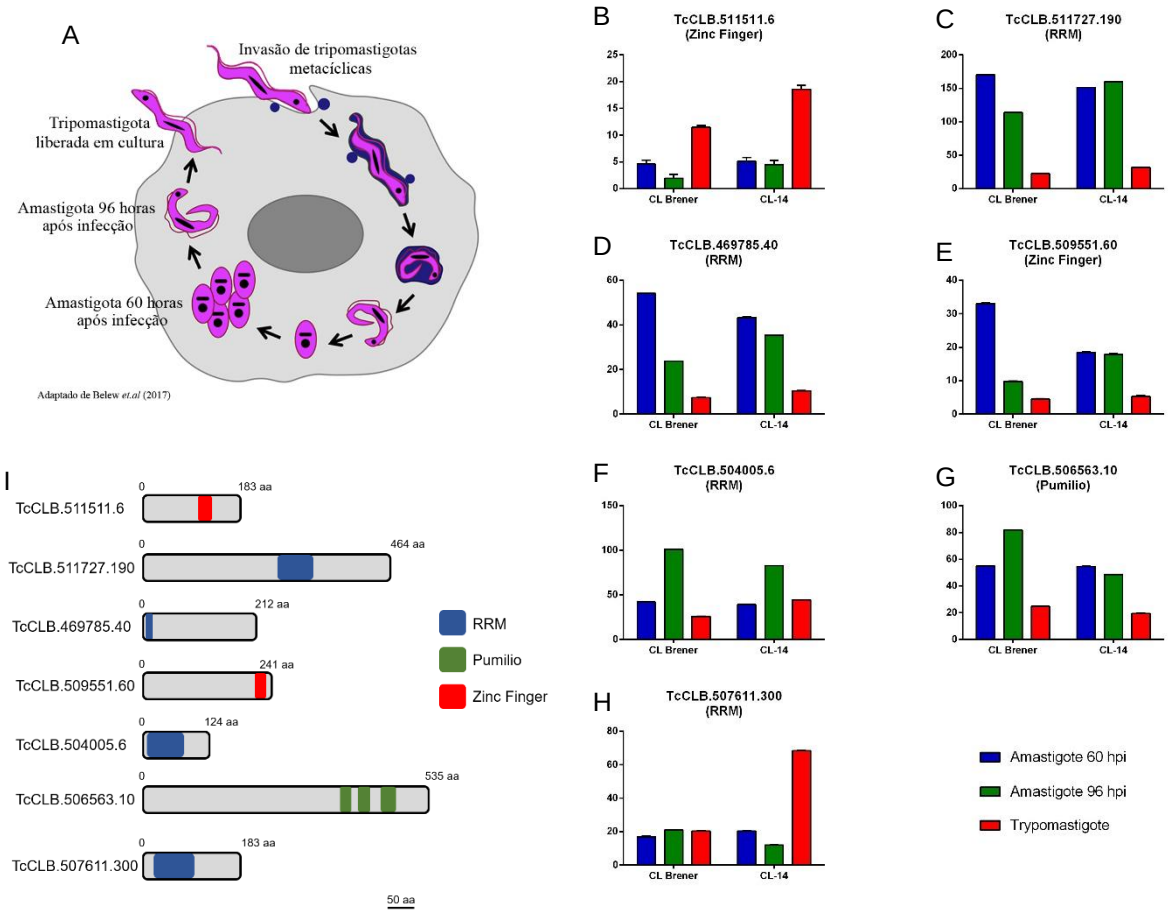


Figura 24: Proteínas de ligação a RNA com expressão aumentada durante o ciclo intracelular do *T. cruzi*. Identificação das proteínas de ligação a RNA diferencialmente expressas entre as formas replicativa e sanguínea. (A) Ciclo intracelular do *T. cruzi* representando os estágios amastigotas com 60 e 96 horas após a infecção e tripomastigotas liberadas em cultura, dos quais foram extraídos mRNAs dos clones CL Brener e CL-14. (B) RBP aumenta em tripomastigota de ambos os clones. (C-G) Proteínas de ligação a RNA diferencialmente expressas em amastigotas nos clones CL Brener e CL-14. (H) Transcrito codificante para uma RBP expressa apenas em tripomastigotas de CL-14. (I) Representação esquemática da sequencia das RBPs com expressão aumentada ao longo do ciclo intracelular do *T. cruzi*.

5.2 Análises *in silico* da TcRBP300

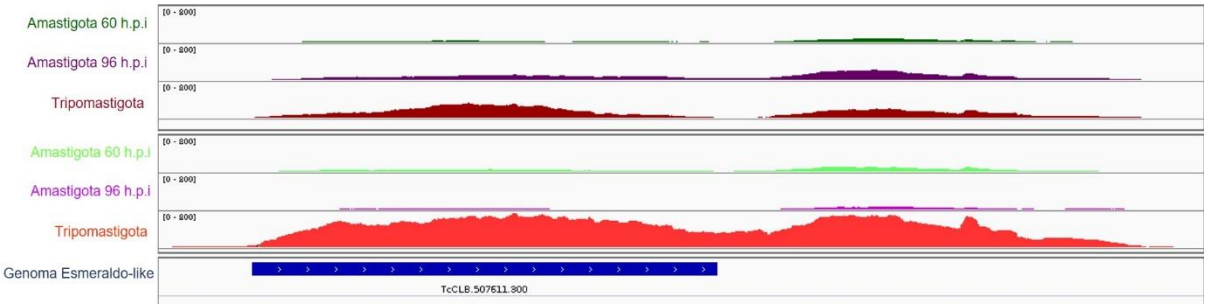
Transcriptoma dos clones virulento e não virulento do *T. cruzi*, a partir de formas amastigota coletadas 60 e 96 horas após a infecção e tripomastigotas liberadas em cultura revelou que o transcrito codificante para a TcRBP300 apresenta expressão diferencial quando os clones CL Brener e CL-14 são comparados. O aumento na expressão da TcRBP300 na forma tripomastigota somente do clone avirulento, CL-14, pode ser observado na figura 25A que mostra o mapeamento das *reads* na região correspondente a esse gene nos estágios amastigota 60 horas após a infecção (verde), amastigota 96 horas após a infecção (roxo) e tripomastigota (vermelho) dos clones CL Brener (cores escuras) e CL-14 (cores claras). Dados depositados no banco de dados TritrypDB mostram que a região codificadora do gene codificante para a proteína de ligação a RNA TcRBP300, que está anotada como "*U1A small nuclear ribonucleoprotein, putative*" e recentemente reconhecida como "*RNA-binding protein 14*", está localizado no cromossomo 37 do alelo Esmeraldo-like do clone CL Brener do *T. cruzi* e é composto por 492 pares de bases. O produto desse gene é uma proteína com 163 aminoácidos que apresenta um domínio de reconhecimento a RNA (RRM) presente entre os nucleotídeos 2 e 78 e ponto isoelétrico básico de 9,53. Novamente, a presença de um domínio RRM juntamente a existência de um pI elevado são evidências que indicam se tratar de uma proteína que interage com RNA. Comparação entre os alelos TcCLB.507611.300 e TcCLB.507723.120, codificantes para essa proteína de ligação a RNA mostram pequenas diferenças na localização cromossômica e no pI (Tabela 8). Os dois alelos codificantes para a TcRBP300 apresentam 99,6% de similaridade na sequencia de nucleotídeos (Figura 25B) e 100% de similaridade na sequencia de aminoácidos (Figura 25C).

Tabela 8: Principais características dos alelos codificantes para a proteína de ligação a RNA com expressão diferencial entre os clones CL Brener e CL-14.

	TcCLB.507611.300	**TcCLB.507723.120**
Alelo	Esmeraldo-like	Não Esmeraldo-like
Localização	Posição 87.471 a 88.002 do cromossomo 37	Posição 87.537 a 88.028 do cromossomo 37
Tamanho do gene	492 bases	492 bases
Tamanho da proteína	163 aminoáciods	163 aminoáciods
Localização do motivo RRM	Entre as bases 2 e 78	Entre as bases 2 e 78
Ponto isoelétrico	9,53	8,78

O alinhamento múltiplo dos genes sintênicos a TcRBP300 presentes nos genomas de outros tripanosomatídoes revelou alta conservação na sequencia de aminoácidos da proteína na região contendo o motivo RRM, localizado na porção N-terminal (Figura 26A). Análises filogenéticas, utilizando as sequencias ortólogas da TcRBP300 retratadas no alinhamento múltiplo, agrupou as sequencias em dois grupos principais, dos Trypanosomas (*T. cruzi* e *T. brucei*) e das Leishmanias (Figura 26B). Não foram encontradas sequencias com homologia a TcRBP300 nos genomas de outros eucariotos.

A



B

```
                1       10        20        30        40        50        60        70        80        90       100       110       120       130
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
Esmeraldo       ATGGAGTTACTGCGAGTTCGAGGACTGGTGTTTTATTTTGGTGTGGAATTTTGTAGATTGACAATACTGCGTTGCTTTAGTGAATATGGTATGCTGTATGACATTCTTTTGGATGTTCCCCGTGGTGAAG
Nao_Esmeraldo   ATGGAGTTACTGCGAGTTCGAGGACTGGTGTTTTATTTTGGTGTGGAATTTTGTAGATTGACAATACTGCGTTGCTTTAGTGAATATGGTATGCTGTATGACATTCTTTTGGATGTTCCCCGTGGTGAAG
Consensus       ATGGAGTTACTGCGAGTTCGAGGACTGGTGTTTTATTTTGGTGTGGAATTTTGTAGATTGACAATACTGCGTTGCTTTAGTGAATATGGTATGCTGTATGACATTCTTTTGGATGTTCCCCGTGGTGAAG

                131     140       150       160       170       180       190       200       210       220       230       240       250       260
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
Esmeraldo       CTCTCGTGTCATATACTGAATCTGCATCTGCTCAGGATGCATATTTAAAAATGAACGGGTTTTTGCTTTTTGGTGAGGCGATAGAGGTAAGTATCACTACAGCTCCAGCTTCGGATATACCCGGGTGCCG
Nao_Esmeraldo   CTCTCGTGTCATATACTGAATCTGCATCTGCTCAGGATGCATATTTAAAAATGAACGGGTTTTTGCTTTTTGGTGAGGCGATAGAGGTAAGTATCACTACAGCTCCAGCTTCGGATATACCCGGATGCCG
Consensus       CTCTCGTGTCATATACTGAATCTGCATCTGCTCAGGATGCATATTTAAAAATGAACGGGTTTTTGCTTTTTGGTGAGGCGATAGAGGTAAGTATCACTACAGCTCCAGCTTCGGATATACCCGGaTGCCG

                261     270       280       290       300       310       320       330       340       350       360       370       380       390
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
Esmeraldo       AGTGACGTATAGACGTTCACAATCTCTGATCCTGCGGGGTGCCTCATCTCTTTGGGTGGAGTTAAATTTGAAACACGTAAAGGGAGTAAATCAAATTGTATTTGTTGATCCAAATACCACAATTGTTTTC
Nao_Esmeraldo   AGTGACGTATAGACGTTCACAATCTCTGATCCTGCGGGGTGCCTCATCTCTTTGGGTGGAGTTAAATTTGAAGCACGTAAAGGGAGTAAATCAAATTGTATTTGTTGATCCAAATACCACAATTGTTTTC
Consensus       AGTGACGTATAGACGTTCACAATCTCTGATCCTGCGGGGTGCCTCATCTCTTTGGGTGGAGTTAAATTTGAAaCACGTAAAGGGAGTAAATCAAATTGTATTTGTTGATCCAAATACCACAATTGTTTTC

                391     400       410       420       430       440       450       460       470       480       490492
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+-|
Esmeraldo       TTTGCAAACCAAAGAATTTCAACAATGATCAAGAAGCTTTTTGATGGACGCATTGATTATAAAGGGAATCCCGTTTCTATCTTATATTTGAAACAGGCTTAG
Nao_Esmeraldo   TTTGCAAACCAAAGAATTTCAACAATGATCAAGAAGCTTTTTGATGGACGCATTGATTATAAAGGGAATCCCGTTTCTATCTTATATTTGAAACAGGCTTAG
Consensus       TTTGCAAACCAAAGAATTTCAACAATGATCAAGAAGCTTTTTGATGGACGCATTGATTATAAAGGGAATCCCGTTTCTATCTTATATTTGAAACAGGCTTAG
```

C

```
                1       10        20        30        40        50        60        70        80        90       100       110       120       130
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
Esmeraldo       MELLRVRGLVFYFGVEFCRLTILRCFSEYGMLYDILLDVPRGEALVSYTESASAQDAYLKMNGFLLFGEAIEVSITTAPASDIPGCRVTYRRSQSLILRGASSLWVELNLKHVKGVNQIVFVDPNTTIVF
Nao_Esmeraldo   MELLRVRGLVFYFGVEFCRLTILRCFSEYGMLYDILLDVPRGEALVSYTESASAQDAYLKMNGFLLFGEAIEVSITTAPASDIPGCRVTYRRSQSLILRGASSLWVELNLKHVKGVNQIVFVDPNTTIVF
Consensus       MELLRVRGLVFYFGVEFCRLTILRCFSEYGMLYDILLDVPRGEALVSYTESASAQDAYLKMNGFLLFGEAIEVSITTAPASDIPGCRVTYRRSQSLILRGASSLWVELNLKHVKGVNQIVFVDPNTTIVF

                131     140       150       160163
                |--------+---------+---------+--|
Esmeraldo       FANQRISTMIKKLFDGRIDYKGNPVSILYLKQA
Nao_Esmeraldo   FANQRISTMIKKLFDGRIDYKGNPVSILYLKQA
Consensus       FANQRISTMIKKLFDGRIDYKGNPVSILYLKQA
```

Figura 25: Expressão TcRBP300 em amastigotas e tripomastigota do *T. cruzi* e alinhamento dos alelos da TcRBP300. (A) Mapeamento da *reads* geradas pelo RNA-Seq dos clones CL Brener (escuro) e CL-14 (claro) a partir de mRNAs de amastigotas coletadas 60 (verde) e 96 (roxo) horas após a infecção e tripomastigotas (vermelho) no cromossomo 37 do genoma Esmeraldo-like do clone CL Brener do *T. cruzi*. A barra azul retrata a região codificadora. (B) Sequencia de nucleotídeos dos alelos Esmeraldo (TcCLB.507611.300) e não Esmeraldo (TcCLB.507723.120) da TcRBP300 foram extraídos do banco de dados TritrypDB e alinhadas usando o programa Multialin (disponível em http://multalin.toulouse.inra.fr/multalin/). (C) Alinhamento entre as sequencias de aminoácidos.

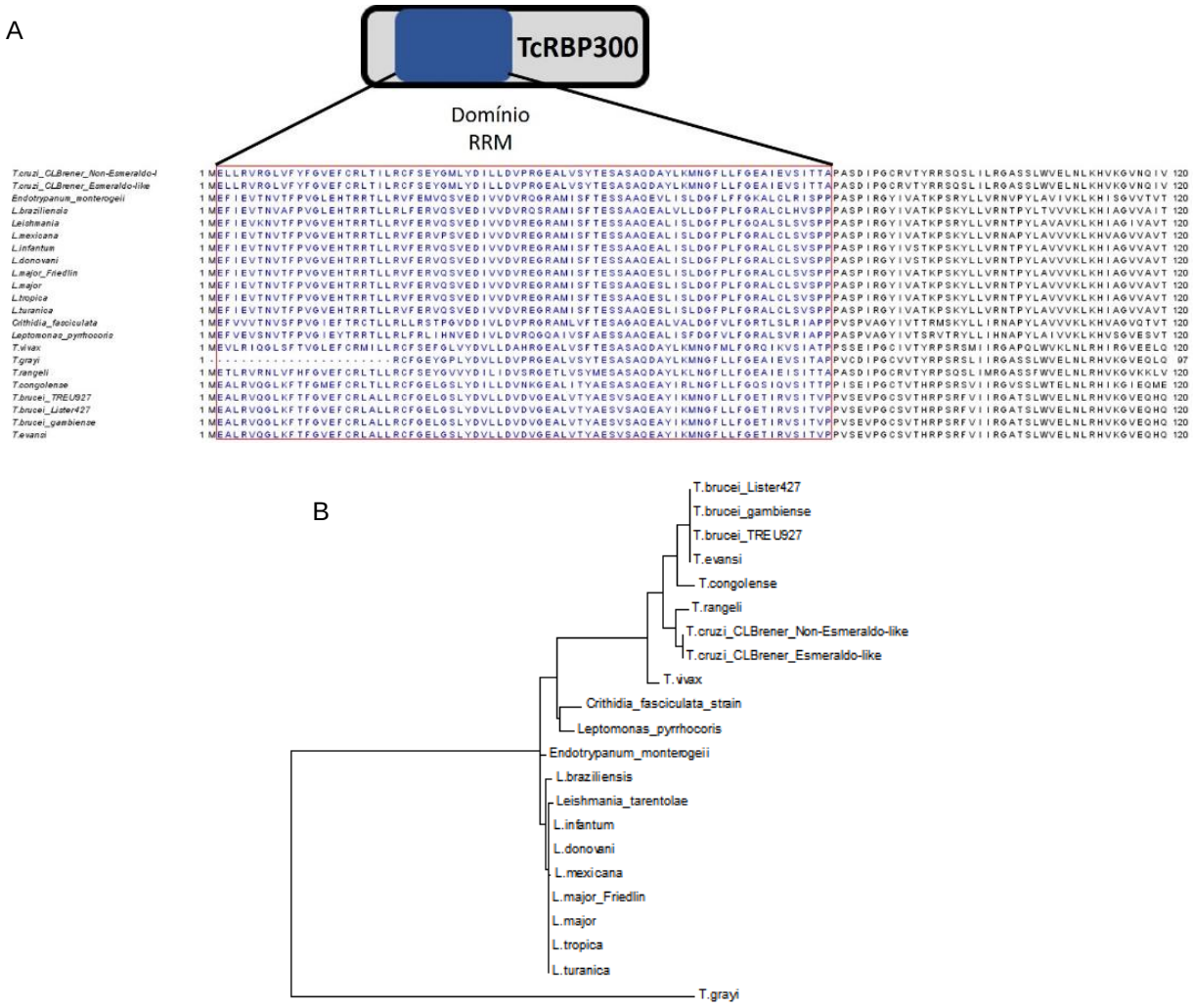


Figura 26: Alinhamento dos genes sintênicos a TcRBP300 e árvore filogenética. Sequencia correspondentes aos ortólogos da TcRBP300 foram extraídos do bando de dados TritryDB. (A) Alinhamento múltiplo dos genes sintênicos a TcRBP300 depositados no banco de dados TritrypDB. (B) Árvore filogenética construída com os genes sintênicos, utizando o software MEGA.

5.3 Geração de linhagens nocautes da TcRBP300

Linhagens nocautes do gene codificante para a TcRBP300 no clone CL-14 ajudariam a compreender os fatores envolvidos com a virulência do *T. cruzi*. A obtenção de linhagens do clone CL-14 nocautes para o gene codificante da TcRBP300 foi realizada utilizando a tecnologia CRISPR/Cas9 pois a endonuclease Cas9, ao gerar uma quebra especifica na dupla fita do DNA na sequencia alvo, estimula o reparo por recombinação homologa (HR), na presença de uma sequencia com homologia à sequencia alvo denominada sequencia doadora. Se a sequencia doadora conter parte do gene alvo interrompido, um gene de resistência a droga ou códons de terminação da tradução, o reparo por HR resulta na interrupção do gene. Para que ocorra a quebra na dupla fita é necessária a presença da proteína Cas9 e de um pequeno RNA completar ao gene alvo, o sgRNA. O gene da endonuclease Cas9 pode ser transcrito e traduzido pelo próprio organismo ou a proteína Cas9 recombinante pode ser fornecida durante a transfecção da célula juntamente com o sgRNAs e as sequencias doadoras. Da mesma maneira, o sgRNA pode ser transcrito pelo organismo ou pode ser sintetizado *in vitro* e entregue ao parasito na transfecção. Para o reparo da quebra na dupla fita, sequencias homólogas ao gene alvo constituídas (i) pelo gene de resistência flanqueado por partes homologas ao gene alvo ou (ii) por um oligonucleotídeo com a sequencia do gene alvo modificada de maneira a conter códons de parada da tradução a fim de gerar uma proteína não funcional, podem ser fornecidas durante a transfecção. O nosso grupo vem testando vários protocolos, e para a geração de linhagens nocaute do gene TcRBP300 duas metodologias que envolvem a tecnologia CRISPR-Cas9 foram utilizadas.

5.3.1 Geração do sgRNA

A identificação das sequencias dos sgRNAs para o gene da TcRBP300 foram obtidas pelo programa EuPaGDT (*Eukaryotic Pathogen CRISPR guide RNA Design Tool*). Esse programa resulta em uma lista de sequencias que podem ser utilizadas para direcionar a clivagem pela nucleasse Cas9 e é recomendável que mais de um sgRNA seja testado. Dentre os sgRNA indicados pelo programa foi escolhido aquele cuja sequencia alvo no gene da TcRBP300 não estivesse na região de recombinação homologa quando o reparo por HR ocorresse com a utilização de uma sequencia doadora contendo partes do gene da TcRBP300 flanquando o gene de resistência à G418. Através de amplificação por PCR (Figura 27A) seguido por uma reação de transcrição *in vitro* pela T7 RNA polimerase (Figura 27B) foi obtido o sgRNA. Esse pequeno RNA contêm a sequência alvo do gene da TcRBP300 seguidos da sequencia PAM de *S. aureus* e de uma sequencia correspondente ao *scaffold* reconhecido pela Cas9 de *S. aureus*.

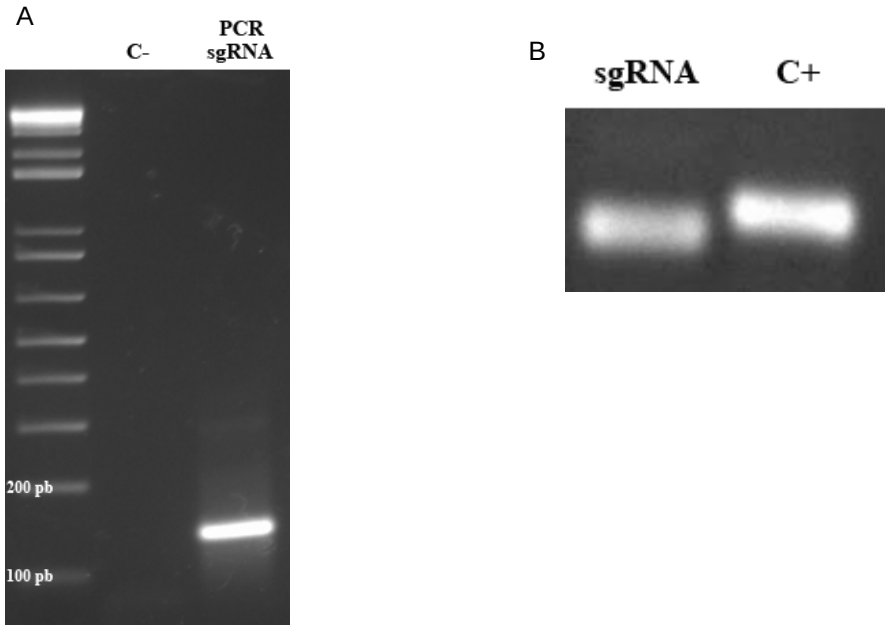


Figura 27: Obtenção do sgRNA alvo para TcRBP300. Escolha do sgRNA foi realizada pelo programa EuPaGDT. (A) Produto de PCR gerado usando iniciadores senso especifico, composto pela sequencia da T7 RNA polimerase e da sequencia alvo da TcRBP300, e anti senso contendo sequencia correspondente ao *scaffold* da Cas9 de *S. aureus*, gerando fragmento de aproximadamente 150 pares de bases. (B) sgRNA após a transcrição *in vitro* do produto de PCR.

5.3.2 Obtenção da sequencia doadora contendo um gene de resistência à droga

A geração de linhagens nocautes para a TcRBP300 contendo resistência à droga, foi realizada interrompendo o gene codificante para a TcRBP300 com o plasmídeo Hx1_Neo_GAPDH. A construção dessa sequencia doadora foi gerada amplificando do DNA genômico do clone CL-14 regiões da TcRBP300 para serem usadas no processo de reparo por recombinação homóloga (Figuras 28A e 28B). Após a clonagem dessas sequencias nas extremidades 5’ e 3’ do plasmídeo Hx1_Neo_GAPDH, foi obtido por PCR a sequencia doadora pKO_TcRBP300_Neo (Figura 28C) contendo regiões para recombinação homóloga (5’KO:TcRBP300 e 3’KO:TcRBP300) e sequencias para o correto processamento do gene de neomicina (5’UTR de Hx1 e 3’UTR de GAPDH) (Figura 28D). Epimastigotas do clone CL-14 foram transfectadas com a proteína recombinante SaCas9, com sgRNA transcrito *in vitro* e o fragmento de 2.670 pares de base contendo a sequencia doadora com o gene de resistência a G418. População resistente a droga G418 foi gerada e clones estão em processo de seleção.

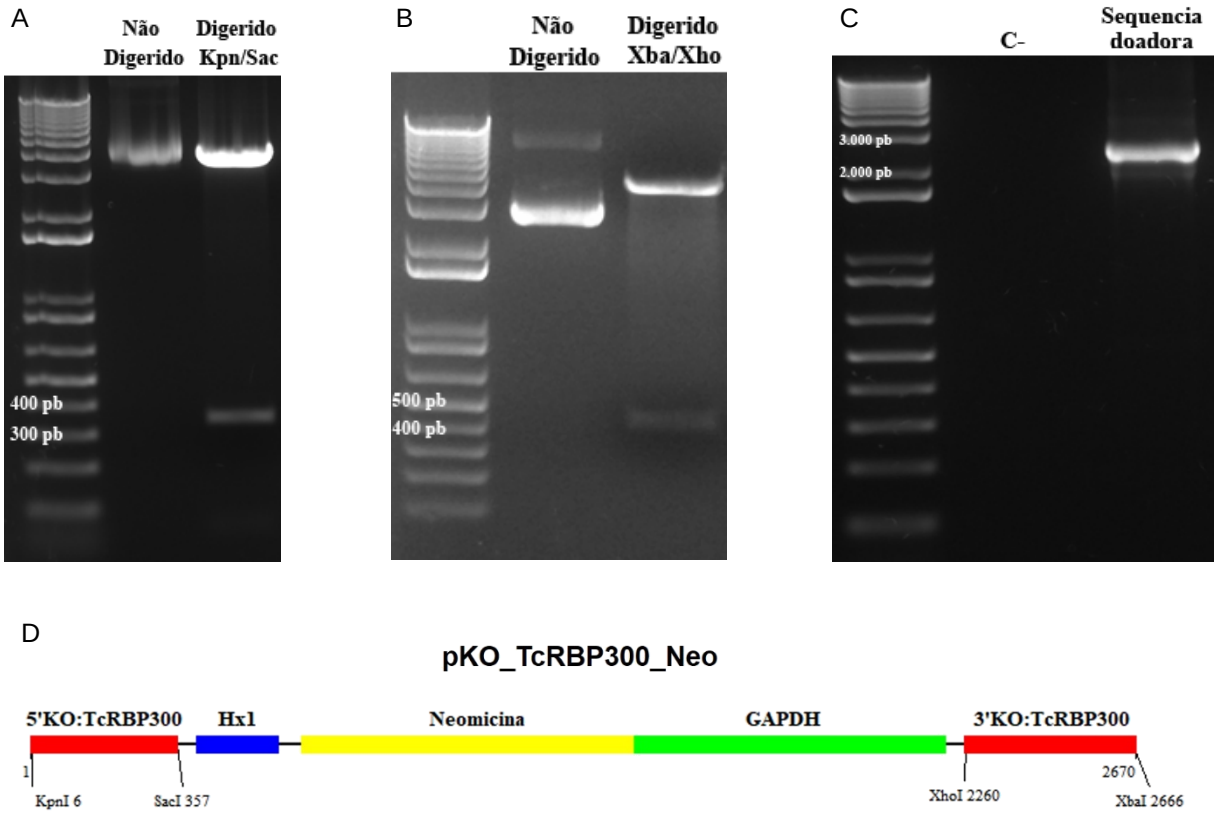


Figura 28: Construção da sequencia doadora com resistência a neomicina para obtenção de linhagens nocautes para a TcRBP300. Interrupção do gene codificante para a TcRBP300 foi realizado com a tecnologia CRISPR/Cas9. Sequencias de 357 e 412 pares de base correspondente as regiões 5' e 3' da TcRBP300, repectivamente foram clonadas no plasmídeo Hx1_Neo_GAPDH. Confirmação das clonagens foi realizada através da digestão do plasmídeo com as enzimas de restrição (A) *Kpn*I e *Sac*I, liberando o fragmento 5'KO:TcRBP300, e (B) *Xba*I e *Xho*I, liberando o fragmento 3'KO:TcRBP300. (C) Amplificação da sequencia doadora por PCR. (D) Representação esquemática da sequencia doadora final contendo as sequencias referentes a 5'UTR de Hx1 (azul), a região codificadora do gene de neomicina (amarelo) e a região 3'UTR de GAPDH (verde), flanqueadas por sequencias homologas a TcRBP300 (vermelho).

5.3.3 Obtenção da sequencia doadora contendo códon de parada de tradução

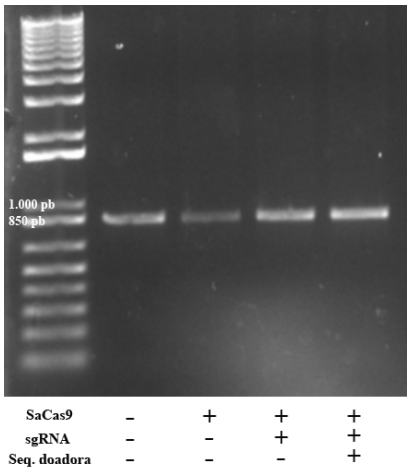
Uma vez que a eficiência de obtenção de nocautes é muito maior quando se utiliza a tecnologia CRISPR-Cas9, torna-se possível obter linhagens nocautes mesmo quando não se utiliza um gene de resistência a droga para selecionar os parasitos após a transfecção. A outra estratégia utilizada para a geração de linhagens nocautes de CL-14 para o gene que codifica TcRBP300 consistiu na transfecção dos parasitos com a endonuclease Cas9, com o sgRNA transcrito *in vitro* e com uma sequencia doadora contendo códons de parada de tradução, a qual interromperia a sequencia da TcRBP300 ao ser utilizada como molde para o reparo por HR. Para isso, foi sintetizado um oligonucleotídeo contendo uma sequencia de 30 nucleotídos homologa à região de quebra da dupla fita, a ser gerada no gene da TcRBP300, seguido por códons de parada de tradução inseridos em três diferentes janelas de leitura (Figura 29A). Para permitir a detecção da incorporação dessa sequencia doadora, um sitio de *Bam*HI foi inserido logo após o último códon de parada. Como não há seleção com droga, a presença desse novo sítio de enzima de restrição permite detectar por análise da digestão do produto de PCR, correspondente a sequencia da TcRBP300, se houve a interrupção do gene. Transfecção de culturas de epimastigotas do clone CL-14 com SaCas9 recombinante, com o sgRNA e o oligonucleotídeo doador ou com apenas SaCas9 utilizada como controle da especificidade do sistema CRISPR/Cas9 foi realizada. Cinco dias após a transfecção, DNA genômico foi extraído a partir da população de parasitos transfectados e a sequencia correspondente ao gene da TcRBP300 foi amplificada (Figura 29B). Em seguida o produto de PCR foi digerido com a enzima de restrição *Bam*HI (Figura 29C). No caso de inserção da sequencia doadora, a digestão com *Bam*HI deveria gerar dois fragmentos de aproximadamente 400 pares de base que não foi observado na população de parasitos transfectados

com todos os componentes do sistema CRISPR/Cas, isto é, com a Cas9 recombinante, o sgRNA e a sequencia doadora.

A

5’GCGATAGAGGTAAGTATCACTACAGCTCCATAGATAGATAGGGATCCGCTTCGGATATACCCGGGTGCCGAGTGACGTATAGACGTT3’

B



C

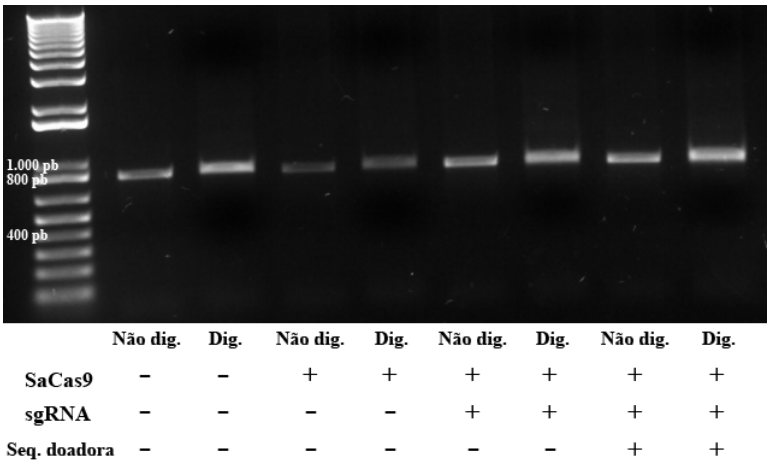


Figura 29: Construção da sequencia doadora com códon de parada de tradução para obtenção de linhagens nocautes para a TcRBP300. Interrupção do gene codificante para a TcRBP300 foi realizado com a tecnologia CRISPR/Cas9. (A) Representação da sequencia doadora contendo 60 bases homólogas a TcRBP300, representada pela cor vermelha, separado pelo códon de parada em três diferentes janelas de leitura, em azul, e pela sequencia correspondente a enzima de restrição *Bam*HI representada em verde. (B) Cinco dias após a transfecção de formas epimastigotas do clone CL-14, DNA genômico foi extraído dos diversos grupos e a sequencia correspondente a TcRBP300 foi amplificada. (C) Produto de PCR foi submetido a digestão com a enzima *Bam*HI. A correta inserção da sequencia doadora deve liberar fragmentos de aproximadamente 400 pares de bases.

5.4 Geração de linhagens super expressando a TcRBP300 e localização celular da TcRBP300

A proteína de ligação a RNA TcRBP300 foi super expressa em formas epimastigotas de CL Brener utilizando o vetor de expressão em *T. cruzi* pROCK. Parasitos do clone CL Brener super expressando essa RBP foram gerados a fim de verificar se o aumento da expressão da TcRBP300 nesse clone afetaria sua capacidade de infeção, uma vez que no clone avirulento CL-14, observamos a expressão aumentada dessa RBP em formas tripomastigotas. Para isso, a região codificadora da TcRBP300 contendo *tag* de HA na sua porção C-terminal foi amplificada a partir do DNA genômico do clone CL-14 e em seguida clonado no vetor pROCK, dando origem ao plasmídeo pROCK_Neo_TcRBP300::HA (Figura 30A). População e clones de parasitos CL Brener transfectados com esse plasmídeo, previamente linearizados com a enzima *Not*I, foram selecionados com a droga G418. Quantidade similares de extratos proteicos da população e de clones super expressando a TcRBP300::HA, foram incubados com anticorpo anti-HA (Figura 30B). O resultado de ensaios de western blot revelou a presença de uma banda com aproximadamente 21 kDa, correspondente ao tamanho da TcRBP300, em variados níveis em todos os clones analisados (Figura 30C). Experimentos para avaliar a infecção dos parasitos do clone CL Brener super expressando a TcRBP300 estão em andamento. A presença da *tag* de HA na região C-terminal de parasitos super expressando a TcRBP300 possibilitou ainda visualizar, por meio de imunofluorescência, que a TcRBP300 encontra-se dispersa por todo o citoplasma de formas epimastigotas (Figura 31).

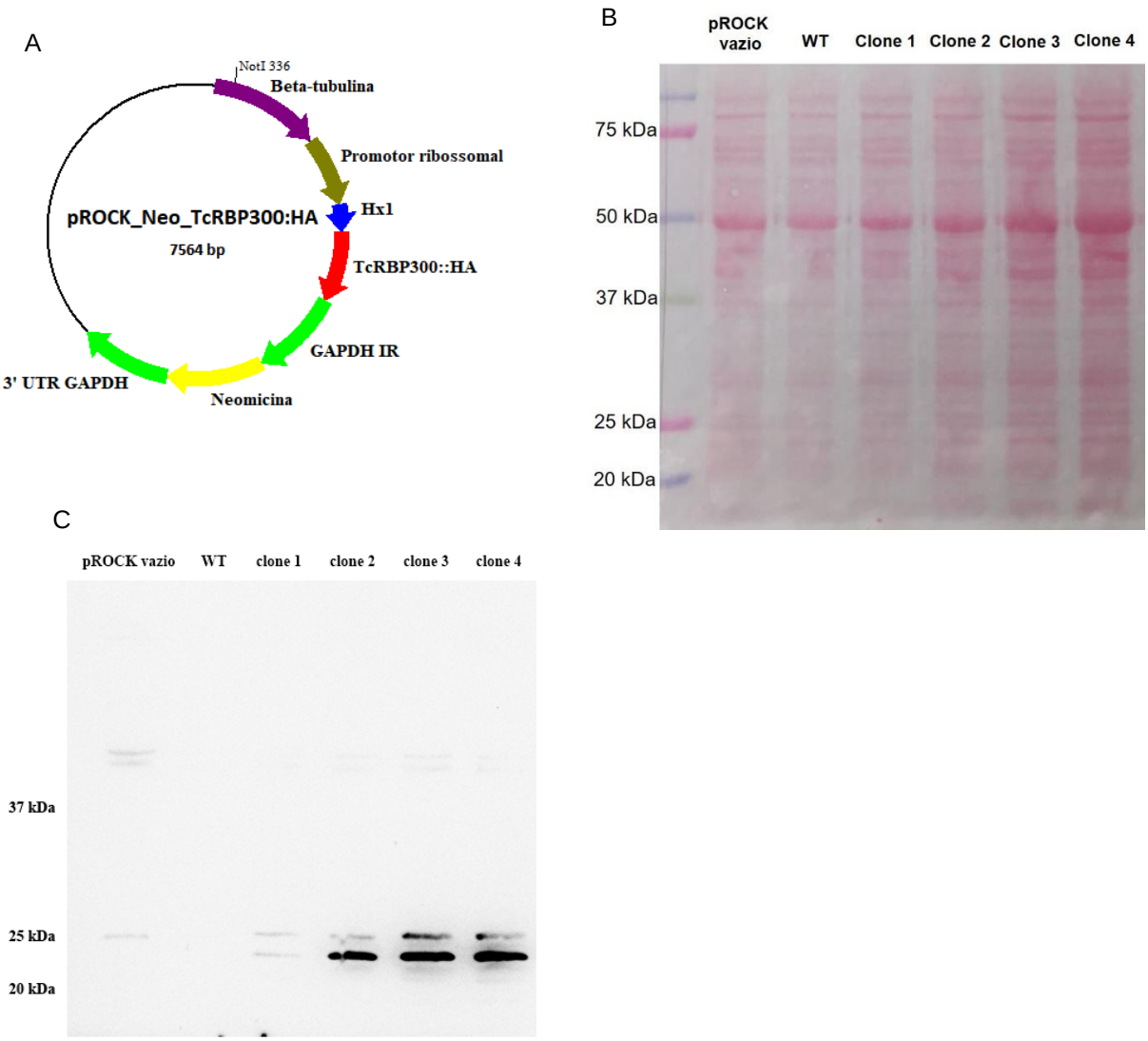


Figura 30: Construção do vetor para super expressar a TcRBP300 e obtenção de linhagens super expressoras. Super expressão do gene da TcRBP300 foi realizado utilizando o vetor de expressão em *T. cruzi* pROCK. (A) Região codificadora da TcRBP300 com *tag* de HA na porção C-terminal foi amplificada do genoma do clone CL-14 e clonado no vetor pROCK, previamente digerido com *Xba*I e *Xho*I para remoção da sequencia de GFP, dando origem ao plasmídeo pROCK_Neo_TcRBP300::HA. Parasitos CL Brener selvagens foram transfectados com esse plasmídeo, previamente linearizado com *Not*I, e população e clones foram selecionados com a droga G418. (B) Membrana corada com Ponceau após a transferência do extrato proteico total de $1x10^6$ parasitos selvagens, parasitos transfectados com o vetor pROCK vazio e quatro clones super expressando a TcRBP300. (C) Após a transferência dos extratos para membrana de nitrocelulose, a membrana foi incuba com anticorpo anti-HA na diluição de 1:5.000 e no dia segunda foi incubada com anticorpo anti-IgG de camundongo diluído 1:10.000.

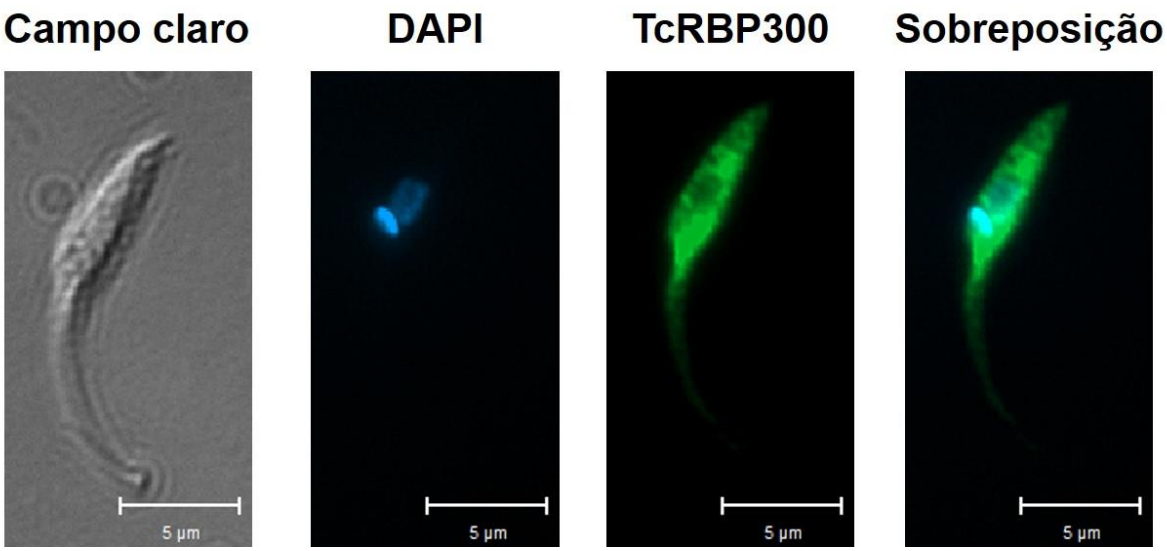


Figura 31: Localização celular da TcRBP300. Formas epimastigotas super expressando a proteína de ligação a RNA, TcRBP300, foram marcadas com anticorpo anti-HA e em seguida com anticorpo anti-rato conjugado com Alexa 488. Núcleo e kinetoplasto foram marcados com DAPI. Imagens foram capturadas no microscópio óptico de fluorescência Axio Imager Z2 – ApoTome 2-Zeiss.

6. DISCUSSÃO

Considerando o fato da transcrição gênica dos tripanosomatídeos ocorrer de forma policistrônica e que as proteínas de ligação ao RNA (RBPs) têm importante papel na regulação dos níveis dos mRNAs no citoplasma, por meio da sua estabilização ou degradação, novas buscas por genes codificadores de proteínas de ligação a RNA presentes no genoma do clone CL Brener do *T. cruzi* (disponível em: http://tritrypdb.org/tritrypdb/) foram conduzidas. Essas análises resultaram na identificação de 253 sequencias contendo domínios de ligação a RNA. Além das 117 sequencias contendo motivo RRM previamente identificadas por Gaudenzi e colaboradores (2005) (37) e Guerra-Slompo e colaboradores (2012) (56), das 96 sequencias contendo o motivo *zinc finger* identificadas por Kramer e colaboradores (2010) (39) e das 16 sequencias contendo o domínio pumilio identificadas por Caro e colaboradores (2006) (41), foram encontradas ainda 24 sequencias contendo domínio de ligação a RNA. Dentre essas sequencias ainda não identificadas estão sequencias contendo o domínio alba, que ainda não haviam sido identificadas e estavam anotadas no banco de dados do genoma do clone CL Brener como genes hipotéticos. É importante ressaltar que o número total de sequencias identificadas em nossas análises é representado por um número menor de genes, uma vez que, devido a natureza híbrida de CL Brener, seus alelos foram anotados separadamente. Desta maneira as 253 sequencias contendo motivos de ligação a RNA representam um total de 147 genes codificadores de RBPs no genoma do clone CL Brener do *T. cruzi*. Dentre as proteínas contendo motivos de ligação a RNA, o motivo RRM é o que está presente em maior quantidade no genoma de *T. cruzi*, sendo esse motivo encontrado em todos os organismos, inclusive em procariotos e nos vírus, porém mais abundante nos eucariotos e o motivo de reconhecimento a RNA mais frequentemente encontrado

nos genomas (82). As 124 sequencias contendo motivo RRM correspondem a 74 genes que é bem próximo dos 75 e 80 genes contendo RRM identificados nos genomas de *T. brucei* e Leishmania respectivamente (37).

Com o sequenciamento dos mRNAs das distintas formas do clone CL Brener do *T. cruzi* foi possível analisar a expressão dos transcritos correspondentes ao grupo de proteínas que tem motivos de ligação ao RNA nas diferentes fases do ciclo de vida deste parasita. Buscas por RBPs diferencialmente expressas entre formas que habitam os diferentes hospedeiros do *T. cruzi*, revelou a presença de cinco proteínas de ligação a RNA com expressão aumentada na forma presente no inseto vetor, isto é, em epimastigotas, comparado às formas amastigota e tripomastigota que são encontradas no hospedeiro vertebrado. É interessante ressaltar que nenhuma dessas cinco RBPs haviam sido previamente descritas na literatura por apresentarem expressão aumentada nas epimastigotas do *T. cruzi*. Dentre esses genes com expressão aumentada no hospedeiro inseto encontram-se os genes TcCLB.511727.270 e TcCLB.511727.290 que codificam RBPs caracterizadas por Umaer e Williams (2015). Esses autores mostraram que essas proteínas estão presentes tanto no núcleo quanto no citoplasma de formas epimastigotas do clone CL Brener e associam-se com o ribossomo através de interação com a unidade 5S do rRNA e com a proteína ribossômica TcL5 (83). Dentre as RBPs com expressão diferencial em epimastigotas a TcCLB.506739.99 destaca-se por ser praticamente expressa apenas em nas formas epimastigotas em comparação com as amastigotas e as tripomastigotas. Assim como as RBPs caracterizadas por Umaer e Williams a TcRBP99 também é exclusiva de tripanosomatídeos, indicando que essas RBPs devem estar envolvidas na regulação pós-transcricional. O estudo dessas proteínas é altamente relevante para se compreender como ocorre o controle da expressão

gênica nesse parasito, uma vez que o processo de transcrição em *T. cruzi* se difere dos outros eucariotos por ser policistrônico e com ausência de promotores conservados para RNA Pol II.

Dados de *microarray* gerados a partir de mRNA proveniente de formas epimastigota, tripomastigota metacíclica, amastigota axênica e tripomastigota liberadas em cultura do clone CL Brener (84), depositados no banco de dados TritryDB, mostram aumento na expressão da TcRBP99 em formas epimastigotas comparado aos outros estágios do *T. cruzi*. Recentemente Li e colaborares também identificaram a presença do gene codificante para a TcRBP99 entre os genes com expressão aumentada nas formas epimastigotas da cepa Y do *T. cruzi* (60). Resultado do mapeamento de *reads* provenientes dos mRNAs de epimastigotas e tripomastigotas metacíclicas da cepa Dm28c no genoma do clone CL Brener (85), depositado no banco de dados TritrypDB, revelou que quantidade maior de *reads* mapearam na região correspondente a TcRBP99 em epimastigotas comparado às tripomastigotas metacíclicas, indicando que na cepa Dm28c ocorre aumento da expressão desse gene em epimastigota comparada à expressão na forma tripomastigota metacíclica (85). No entanto, os dados disponíveis no banco de dados TritrypDB não mostra a presença de um gene ortólogo a TcRBP99 na região sintenica do genoma da cepa Dm28c. Busca por homologia a essa RBP no genoma da cepa Dm28c, através de blast, não resultou em nenhuma sequencia com homologia a TcRBP99, provavelmente porque algumas regiões não foram completamente sequenciadas e a montagem do genoma de Dm28c é muito incompleta (19). Assim, apesar ter sido publicado o transcriptoma de diferentes formas do ciclo de vida do clone Dm28c, não foi possível determinar os níveis de expressão do ortólogo da TcRBP99 nessa cepa com base nos dados gerados por Berna e colaboradores (86).

Apesar de ser altamente expressa em formas epimastigotas, o gene codificante para a proteína de ligação a RNA, TcRBP99, não é um gene essencial uma vez que linhagens nocautes para esse gene foram viáveis. Ao contrário do observado por Ouna e colaboradores (2012) que estudaram o gene ortólogo a TcRBP99 em *T. brucei*, a ausência da TcRBP99 afeta o crescimento de epimastigotas que tem o número de parasitos reduzido no final da fase logaritmica da curva de crescimento, em meio LIT, quando comparado aos parasitos selvagens. Linhagens nocautes mostraram ainda que essa RBP está envolvida no processo de diferenciação de epimastigotas para tripomastigotas metacíclicas que ocorre no hospedeiro invertebrado, uma vez que os resultados mostraram que a taxa de diferenciação foi significativamente aumentada nos nocautes comparada a parasitos selvagens. Por outro lado, aumento na expressão da TcRBP99, cujo mRNA já possui níveis elevados em epimastigotas, também afeta o crescimento e a diferenciação de forma contrária aos nocautes: uma linhagem super expressando a TcRBP99 apresentou aumento na densidade de parasitos comparado à densidade dos parasitos selvagens no nono dia da curva de crescimento e uma taxa de diferenciação em tripomastigotas metacíclicas aproximadamente 50% menor do que a taxa de metaciclogênese dos parasitos selvagens. Recentemente, Alcantara e colaboradores (2018) mostraram que a proteína de ligação a RNA, TcZC3H31 que possui três domínios zinc finger, é essencial para a diferenciação de epimastigotas em tripomastigotas metacíclicas, uma vez que epimastigotas nocautes para essa RBP são incapazes de diferenciarem-se em tripomastigotas metacíclicas (54). Em conjunto, esses dados realizados *in vitro* sugerem que a TcRBP99 deve ligar-se a mRNAs que regulam negativamente a metaciclogênese e em transcritos responsáveis pelo controle da proliferação de epimastigotas. Ensaios de infecção *in vivo*, isto é, infecção de triatomíneos com

epimastigotas nocautes para a TcRBP99 e super expressando essa RBP, irão mostrar se essas diferentes linhagens são capazes de colonizar o intestino do barbeiro, bem como reforçar a hipótese de que a TcRBP99 tem papel na estabilização de mRNAs envolvidos com o crescimento de epimastigotas e no controle da metaciclogênese.

A re-expressão do gene codificante para a TcRBP99 em linhagens nocautes restaurou o fenótipo de crescimento e diferenciação se comportando como parasitos selvagens. Apesar dos níveis de expressão do mRNA referente a essa RBP nas linhagens re-expressores estarem próximos ao selvagem, os clones re-expressores não foram capazes de restabelecer totalmente o crescimento, sendo este significantemente menor comparado ao crescimento dos parasitos selvagens e dos parasitos transfectados com o pROCK vazio. Este fato pode ser explicado pelo processamento do gene da TcRBP99 nas diferentes linhagens usadas no decorrer deste trabalho. Nas linhagens re-expressoras a região codificadora correspondente ao gene da TcRBP99 está flanqueada pelas regiões 5'UTR e 3'UTR dos genes Hx1 e GAPDH, respectivamente, que juntas conseguem aumentar a expressão de um gene (61). Por outro, lado os parasitos selvagens e os transfectados com o vetor pROCK vazio, possuem as regiões 5'UTR e 3'UTR do gene da TcRBP99, indicando que as regiões UTRs da TcRBP99 devem possuir elementos importantes para aumentar a estabilidade do mRNA codificante para a TcRBP99 no citoplasma, para que ele esteja disponível para tradução. Com isso o mRNA codificante para a TcRBP99 estará disponível por mais tempo para que a proteína da TcRBP99 seja sintetizada e então interaja com os mRNAs envolvidos no crescimento e na diferenciação de formas epimastigotas.

A hipótese de que TcRBP99 teria um papel na estabilização dos mRNAs com os quais interage se baseia nas análises de sequenciamento dos mRNAs expressos nas

formas epimastigotas de linhagens nocautes e de parasitos selvagens que mostraram que a ausência da TcRBP99 resulta na redução de expressão de vários genes. Com o sequenciamento foram geradas aproximadamente 3.500.000 *reads* para cada biblioteca que está dentro do esperado considerando a plataforma utilizada no sequenciamento, MiSeq, e ainda que sete bibliotecas foram sequenciadas (disponível em: https://www.illumina.com/systems/sequencing-platforms/miseq/specifications.html). Como os dados do transcriptoma das linhagens nocautes mostraram alteração na expressão de transcritos que possuem normalmente expressão aumentada em formas epimastigota comparado as formas amastigotas e tripomastigotas, pode-se inferir que a expressão aumentada da TcRBP99 em epimastigotas regula positivamente a expressão dos mRNAS essenciais para a proliferação do parasito no estágio epimastigota.

Entre os genes que tiveram alteração significativa nos níveis de expressão na ausência da TcRBP99 está um gene que codifica para um membro da família dos transportadores de aminoácido (TcCLB.506153.10). Esse gene, bem como os outros 16 membros da família, que apresentam expressão aumenta nas formas epimastigotas em relação a amastigotas e tripomastigotas liberada em cultura, também se encontra entre os genes com expressão aumentada em epimastigotas da cepa Y (60). A presença de gene codificante para um transportador de aminoácidos, entre os genes diferencialemente expressos entre parasitos nocautes para TcRBP99 e os selvagens, está de acordo com a hipótese de que a TcRBP99 regula positivamente a expressão de genes necessários para a proliferação de epimastigotas, uma vez que a principal fonte de energia de epimastigotas ocorre através do metabolismo de aminoácidos (87). Tendo em vista que parasitos nocautes apresentaram um crescimento reduzido em comparação com o crescimento dos

parasitos selvagens, pode-se explicar esse fenótipo como consequência da redução nos níveis de expressão do gene que codifica um transportador de aminoácido. Na ausência de transportadores não há entrada de aminoácidos o que acarreta na redução do metabolismo e consequentemente do crescimento. Observando que a deleção de apenas um alelo da TcRBP99 já reduz pela metade os níveis de mRNAs codificantes para transportadores de aminoácidos, podemos inferir que a TcRBP99 deve regular praticamente todos os membros da família dos transportadores de aminoácidos composta de 17 membros que tem expressão aumenta em epimastigotas em comparação a amastigotas e tripomastigotas liberadas em cultura.

Além do transportador de aminoácido, um transcrito que codifica para uma proteína associada a diferenciação (PAD), TcCLB.506551.10, encontra-se entre os genes que tiveram expressão reduzida nos parasitos nocautes comparada a expressão de parasitos selvagens. Esse gene faz parte de uma família de genes codificantes para a família das PADs e que apresentam expressão aumentada em epimastigotas de CL Brener, como revelou o *heatmap* gerado pelos dados de RNA-seq proveniente das formas epimastigota, amastigota e tripomastigota liberada em cultura. Expressão aumentada em epimastigotas dos genes de PAD também foram detectadas nas cepas Y e Dm28c, quando analisamos as contagens de *reads* geradas e mapeadas no genoma do clone CL Brener (60, 85). Em *T. brucei* as proteínas associadas a diferenciação estão organizadas em sequência no final de uma unidade policistrônica e possuem um domínio conservado denominado *nodulin-like* (88). Dean e colaboradores (2009) caracterizam dois membros da PAD de *T. brucei*, denominados PAD1 e PAD2, que apresentam expressão aumentada nas formas *stumpy* e procíclica, respectivamente, e mostraram que a expressão da PAD2, similar ao gene TcCLB.506551.10 de *T. cruzi*, é fortemente induzida durante a diferenciação

da forma sanguínea *stumpy* para a forma procíclica presente no inseto. A depleção de mRNA correspondente a PAD2 em *T. brucei* reduz a diferenciação de tripomastigotas sanguíneas em formas procíclicas que ocorrem na presença de aconitato. Os dados sobre os ortólogos desse gene em *T. brucei* sugerem que as PADs, ou proteínas associadas a diferenciação atuam como transdutores de sinal de diferenciação. Com base nesses dados podemos especular que no *T. cruzi* a expressão das proteínas associadas a diferenciação é induzida durante o processo de diferenciação de tripomastigotas sanguíneas em epimastigotas, durante o hematofagismo, sendo proteínas importantes para a manutenção das formas epimastigotas no inseto vetor. De acordo com essa hipótese, a redução nos níveis de expressão de membros dessa família, como ocorreu nos parasitos nocaute da TcRBP99, fizeram com que as formas epimastigotas diminuissem a sua capacidade de proliferação e com isso, houve a sinalização para a diferenciação em tripomastigotas metacíclicas. Amplificação do domínio conservado *nodulin-like* por PCR em tempo real mostrou que houve redução nos níveis de expressão do mRNA de genes que possuem esse domínio, como as PADs, nas linhagens nocautes, indicando que a TCRBP99 deve interagir com os membros da PAD. Ensaios de co-imunoprecipitação de RNA também confirmaram a interação entre a TcRBP99 e o mRNA codificante para a PAD que tem expressão diferencial entre epimastigotas selvagens e nocaute para a TcRBP99 de *T. cruzi*.

Dentre os genes que tem expressão reduzida em linhagens nocautes para TcRBP99 estão três genes codificantes para os retrotransposons hot spot (RHS) que são a quarta família multigênica com mais membros no genoma de *T. cruzi* e agrupam-se em regiões subtelomêricas (14, 89). Estudo da família RHS em *T. brucei* dividiu os membros dessa família em seis subfamílias (RHS1 a RHS6) com base no domínio C-

terminal e mostrou que a identidade dos genes que compõe essa família varia de 13% a 96% (90). Análises de *heatmap* com os dados do transcriptoma do clone CL Brener mostram que a expressão dos membros da família RHS é aumentada em epimastigotas. Recentemente, estudo da proteína nuclear de ligação a RNA, denominada TbRRM1, pelo grupo de pesquisa de Isabel Roditi (2015) que mostrou que na forma procíclica a TbRRM1 interage com mRNAs codificantes para a família dos RHS bem como com as proteínas RHS e as histonas (91). A TbRRM1 que é expressa constitutivamente nas formas procíclica e sanguínea do *T. brucei* e contêm três motivos RRM e dois motivos *zinc finger* (92), associa-se com genes e proteínas RHS reconhecendo alguns transcritos específicos para retê-los no núcleo e evitar que ocorra a tradução (91). Em discordância com os dados de RNA-seq, as análises de PCR em tempo real mostraram que não houve alteração na expressão global dos genes codificantes para os membros da RHS nos parasitos nocautes e hemi-nocautes comparado as linhagens selvagens. Da mesma maneira não foi observada diferença entre os níveis de expressão das RHS nos parasitos re-expressando a TcRBP99 e os selvagens. Algumas explicações para essa discordância entre os dados de RNA-seq e de PCR em tempo real podem ser levantadas, como por exemplo: 1- que a deleção do gene codificante para a TcRBP99 não afeta a expressão de todos os membros da família RHS e quando analisada a expressão por PCR em tempo real amplifica-se todos os transcritos dessa família, uma vez que os iniciadores usados anelam em região conservada para membros da RHS; 2- o aparecimento desses específicos membros da RHS entre os genes diferencialmente expressos pode ter ocorrido devido a um erro de mapeamento, já que as famílias multigênicas apresentam sequencias conservadas o que dificulta o mapeamento das *reads* nos diversos membros pois uma mesma *read* pode alinhar em diversos membros ou até mesmo ocorrer um

empilhamento de *reads* em um único membro, sendo difícil identificar de que membro aquela *read* é proveniente.

Em *T. brucei* o gene sintênico (Tb927.5.1570) a TcRBP99 apresenta níveis similares de expressão nas formas procíclicas, presentes na mosca tsé-tsé e que correspondem as epimastigotas em *T. cruzi*, e sanguínea, correspondente as tripomastigotas sanguíneas de *T. cruzi* presente no hospedeiro vertebrado (93, 94). Ouna e colaboradores (2012) mostraram por meio da técnica de RNAi que a redução da expressão do gene Tb927.5.1570, codificante para a ZC3H12, não afetou o crescimento das formas procíclicas. Ao contrário da TcRBP99 de *T.* cruz*i*, que estabiliza os mRNAs aos quais interage, a RBP de *T. brucei* codificada pelo gene ZC3H12 interage com a exoribonuclease XRNA que está envolvida na via de degradação dos mRNAs, indicando que essa RBP recruta a XRNA levando a desestabilização dos mRNAs aos quais ela se liga (94). Apesar dos dados de sintenia indicarem que o gene que codifica a TcRBP99 seria ortólogo ao gene ZC3H12 de *T. brucei*, que codifica uma proteína envolvida na desestabilização de mRNAs, os dados de linhagens nocaute para a TcRBP99 indicaram que, no *T. cruzi*, a TcRBP99 é uma proteína que estabiliza os mRNAs aos quais ela interage. Alinhamento entre as sequencias de aminoácidos correspondentes as proteínas TcRBP99 e a ZC3H12 revela similaridade apenas nos 45 primeiros aminoácidos, que fazem parte da região N-terminal onde está localizado o motivo *zinc finger* (Figura 32). Essa baixa similaridade poderia indicar que a presença do domínio *zinc finger*, importante na interação da proteína com o mRNA, não está envolvido diretamente com o papel dessas RBPs na estabilização ou degradação dos mRNAs com os quais elas interagem. Com base nesse alinhamento e nos dados de Ouna e colaboradores (2012), podemos especular que o restante da proteína, que não contem o domínio de

ligação a RNA, poderia estar envolvido em interações com outras proteínas e que resultam na estabilização ou degradação do mRNA. Ensaios de co-imunuprecipitação capazes de identificar as proteínas com as quais TcRBP99 interage, em formas epimastigotas de *T. cruzi*, poderiam desvendar o mecanismo de ação bem como a função dessa proteína. Por outro lado, não devemos descartar a possibilidade de que houve um erro durante a montagem do gene codificante para essa RBP do *T. cruzi* (TcCLB.306739.99) ou do *T. brucei* (Tb927.5.1570), uma vez que a região dos genomas desses dois tripanosomatídeos, onde esse gene está localizado, são sintênicas (Figura 33).

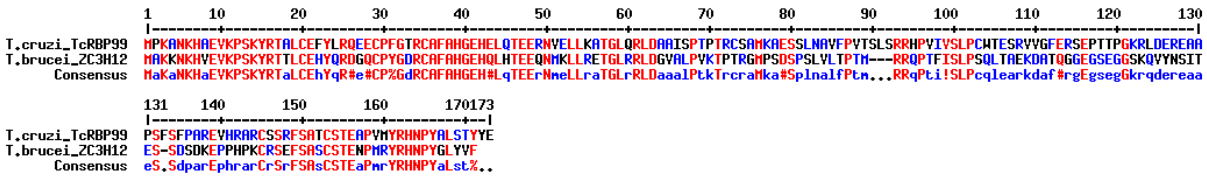


Figura 32: Alinhamento das sequencias da TcRBP99 e ZC3H12. Sequencia de aminoácidos correspondentes as RBPs TcRBP99 (TcCLB.506739.99) e seu ortólogo em *T. brucei* ZC3H12 (Tb927.5.1570) foram extraídos do banco de dados TritrypDB e alinhadas usando o programa Multialin (disponível em http://multalin.toulouse.inra.fr/multalin/).

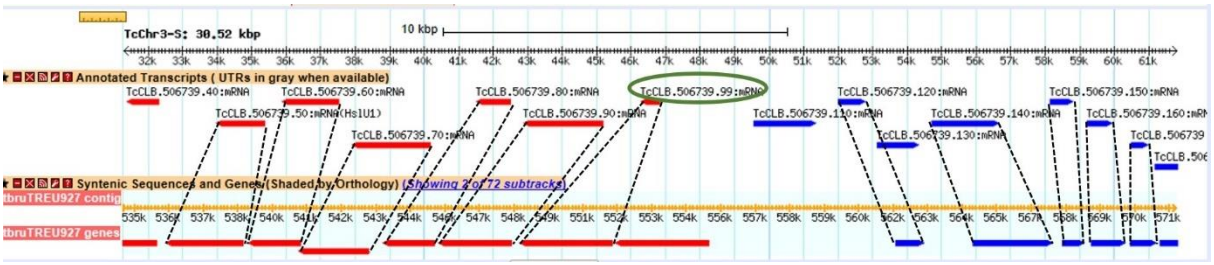


Figura 33: Região no cromossomo 3 do *T. cruzi* onde está localizada o gene da TcRBP99 e seu correspondente em *T. brucei*. Região do cromossomo 3 do alelo Esmeraldo onde está localizada a TcRBP99, destacada pelo círculo verde, e a região correspondente no genoma do *T. brucei* foram extraídas do bando de dados TrytripDB. Barras vermelhas e azuis representam genes presentes nessa região e traços pontilhados pretos indicam genes sintênicos nessa região do genoma do *T. cruzi* (parte superior) e o *T. brucei* (parte inferior). Adaptado do bando de dados TritrypDB.

Procurando identificar sítios de ligação da TcRBP99 nos seus possíveis mRNAs alvos, foram feitas pesquisas por motivos conservados nos transcritos que tiveram seus níveis de expressão alterados nas linhagens onde o gene dessa RBP foi deletado. Nas regiões 3'UTR dos transcritos correspondentes as proteínas hipotéticas, a PAD e o transportador de aminoácidos as bases A e C são mais frequentes, enquanto que os mRNAs codificantes para membros da família RHS apresentam, em suas regiões 3' UTR, sequencias conservadas contendo predominantemente as bases G e U. Mörking e colaboradores mostraram que a proteína de ligação a RNA ZFP2 liga-se a sequencias ricas em A (53), enquanto que a ZFP1 reconhece sequencias ricas em C (95). Anteriormente Pérez-Dias e colaboradores (2007) já haviam demonstrado que a TcRBP19, que possui o motivo RRM, interage preferencialmente com regiões ricas em C. Considerando que nenhum desses três trabalhos mostraram papel desestabilizador das RBPs estudadas e que elementos ricos em A e U (ARE) são conhecidos por serem desestabilizadores, ou seja, reconhecidos por proteínas que estimulam a degradação do mRNA, (96) a ideia de que a TcRBP99 atua estabilizando os mRNAs codificantes para a proteína associada a diferenciação e o transportador de aminoácidos é reforçada, já que nas regiões 3'UTR desses transcritos não foi identificadas predominância das bases A e U. Por meio de ensaios de CLIP-Seq será possível identificar em quais sequencias, presentes nas regiões 3'UTR dos transcritos, a TcRBP99 interage.

Na segunda parte desse trabalho procuramos por proteínas de ligação a RNA que poderiam explicar a diferença de virulência apresenta pelos clones CL Brener e CL-14. Análise do transcriptoma das formas amastigotas e tripomastigotas, que fazem parte do ciclo intracelular do parasito e que estão envolvias com os fatores de virulência, de ambos os clones revelou a presença de 7 genes codificantes para

proteínas de ligação a RNA diferencialmente expressas nessas formas. Seis delas apresentam expressão diferencial entre os estágios replicativo e infectivo do hospedeiro vertebrado e apenas um gene codificante para RBP é diferencialmente expresso entre os clones CL Brener e CL-14. Enquanto o transcrito codificante para a TcCLB.507611.300 apresenta expressão constitutiva no clone CL Brener, em CL-14 ocorre um aumento significativo da expressão desse gene nas formas tripomastigotas. Os dados de RNA-seq mostraram haver uma redução na expressão gênica de diversas famílias codificadoras de proteínas de superfície em tripomastigotas do clone CL-14 (avirulento) comparadao com CL Brener (24). Com base nesse dado e na grande diferença observada no padrão de expressão da TcRBP300, o aumento da expressão dessa RBP em tripomastigota de CL-14 poderia atuar desestabilizando os mRNAs envolvidos com a virulência do *T. cruzi* e que tem a expressão aumentada na forma tripomastigota. Estudos com o ortólogo da TcRBP300, altamente conservado entre o tripanosomatídeos, em *T. brucei* indicam que se trata da ribonucleoproteína U1A, um componente do spliceossoma (97). Tendo em vista que a U1 é importante para a junção *cis* e não para o *trans-splicing* e ainda que o *cis splicing* é um evento raro em tripanosomatídeos, a presença do transcrito codificante para a proteína ortóloga à ribonucleoproteína U1A entre os genes diferencialmente expressos não era esperado. Podemos especular que em *T. cruzi* essa proteína não esta relacionada com o processo de *cis splicing*, mas que o aumento dessa RBP, apenas em tripomastigotas de CL-14, poderia estar envolvido na desestabilização de mRNAs envolvidos com a virulência do *T. cruzi*. Linhagens de parasitos super expressando a TcRBP300 com *tag* de HA no clone CL Brener, poderia ajudar a identificar se ocorre interação dessa RBP com mRNAs codificantes para proteínas de superfície que apresentam expressão aumentada em tripomastigotas. Além disso essas linhagens

super-expressoras permitirão analisar se o aumento da expressão dessa RBP em tripomastigotas de CL Brener altera a capacidade das tripomastigotas infectarem células, bem como a multiplicação das amastigotas dentro das células e a liberação de tripomastigotas no sobrenadante da cultura. Análise da presença da TcRBP300 com *tag* de HA, por western blot, mostrou que os clones super expressores, expressam a TcRBP300::HA em níveis variados.

Tendo em vista a eficácia do sistema CRISPR/Cas9 na edição de genes exógenos, endógenas e até com múltiplas cópias no genoma do *T. cruzi* (68, 69, 98), a tecnologia CRISPR/Cas9 foi utilizada para gerar linhagens de CL-14 nocautes para a TcRBP300. Ensaios de infecção *in vitro* e *in vivo* com parasitos nocautes para a TcRBP300 revelariam se essa RBP é imporante para a virulência do parasito, e então análise da expressão gênica global de linhagens nocautes e selvagens ajudariam a identificar os genes envolvidos com a virulência do *T. cruzi*. Transfecção de linhagens selvagens do clone CL-14 com sgRNA que tem como alvo a TcRBP300, complexado com a proteína recombinante Cas9 de *Staphylococcus aureus* e dois fragmentso de DNA para serem utilizados como sequencias doadoras para o reparo por recombinação homóloga, foi a estratégia escolhida para obtenção de nocautes. Diferente do nocaute gerado pela metodologia convencional, usado na deleção do gene que codifica a TcRBP99, a tecnologia de CRISPR/Cas9 para o nocaute do gene da TcRBP300 foi escolhido com base nos trabalhos de Soares Medeiros e colaboradores (2017) e os dados do nosso grupo (Burle-Caldas et al., submetido) que mostraram que o uso da proteína recombinante SaCas9 resulta em alta eficiência de edição gênica em *T. cruzi* e em outros tripanosomatídeos (99). Como, de acordo com a nossa estratégia, a sequencia doadora adiciona um códon de parada de tradução e uma sequencia correspondente a enzima de restrição *Bam*HI na sequencia da

TcRBP300, a presença de parasitos nocautes para o gene alvo na população transfectada pode ser avaliada pela digestão do produto de PCR, correspondente a amplificação desse gene, com a enzima *Bam*HI. Entretantoo resultado dessa análise mostrou que não houve clivagem do produto de PCR, indicando que a inserção da sequencia doadora durante o reparo da quebra de dupla fita por recombinação homóloga não aconteceu com uma eficiência minimamente detectável. Entretanto, presença de parasitos nocautes para a TcRBP300 na população transfectada ainda não deve ser descartada, pois poderia ter ocorrido a edição gênica em uma quantidade parasitos muito pequena que estaria representada por um número muito reduzido de fragmentos de DNA amplificados a partir dessa população. Nesse caso, a clonagem de parasitos e análises por PCR e digestão de um grande número de clones ou mesmo a retransfecção da mesma população com o complexo Cas9 e sgRNA e a sequencia doadora poderiam resultar na identificação de linhagens nocautes. Como, paralelamente, utilizamos uma estratégia alternativa na qual a sequencia doadora utilizada na transfecção contem gene de resistência a neomicina e parasitos resistentes a droga foram selecionados, podemos supor que na população resistente poderemos encontrar parasitos com a deleção de um ou até mesmo dos dois alelos da TcRBP300. Análises dos clones que estão em andamento irão permitir afirmar se com a tecnologia de CRISPR/Cas9 foi possível interromper os dois alelos do gene de TcRBP300 e obter linhagens nocautes do clone CL-14. Uma vez que essas linhagens nocauates forem obtidas, experimentos de infecção celular e RNA-Seq das linhagens nocautes serão realizados com o objetivo de verificar se há alteração no ciclo intracelular das linhagens nocautes e para identicar os genes que apresentam alteração de expressão na ausência da TcRBP300.

Ambas as proteínas de ligação a RNA estudadas estão dispersas no citoplasma de epimastigotas, o que era esperado uma vez que o citoplasma é o local onde há acúmulo dos mRNAs. Por microscopia foi possível observar ainda que em alguns

locais o sinal da fluorescência é mais intenso, podendo essas regiões corresponder a grânulos de stress, *P-bodies* ou grânulos citoplasmáticos, que são conhecidos por serem locais de armazenamento de mRNA (45). Ensaios de imunofluorescência estão sendo planejados para verificar se ocorre a co-localização da TcRBP99 com componentes dos grânulos citoplasmáticos. Uma vez que esses grânulos foram descritos por estarem envolvidos na estabilização dos mRNAs (100), se for verificada a co-localização da TcRBP99 com essas estruturas celulares poderíamos especular que uma possível função dessa RBP seria a de conduzir os mRNAs até esses grânulos para armazená-los.

7. CONCLUSÃO

O peculiar processo de transcrição gênica dos tripanosomatídeos conduz à ideia de que a regulação gênica ocorre após a geração dos mRNAs maduros, que se dá principalmente pela interação das proteínas de ligação a RNA com os mRNAs. A expressão dos genes codificantes para proteínas de ligação a RNA nos diversos estágios do ciclo celular de diferentes cepas do *T. cruzi*, revelou que a TcRBP99 é específica de epimastigotas de CL Brener e que a TcRBP300 apresenta um padrão de expressão em tripomastigotas do clone CL-14 (avirulento) distinto do padrão de expressão do clone CL Brener (virulento).

O gene codificante para a TcRBP99 não é essencial para a sobrevivência de epimastigotas cultivadas *in vitro*. A TcRBP99 é uma proteína citoplasmática que aparentemente interage com mRNAs que possuem expressão aumentada em epimastigotas e atua, possivelmente, como uma proteína estabilizadora desses transcritos (Figura 34). Entre os transcritos que tem a expressão controlada por essa RBP alguns codificam proteínas envolvidas com o transporte de aminoácidos para que esses possam ser disponibilizados para o metabolismo das epimastigotas no inseto vetor. Com isso os epimastigotas são capazes de proliferar, por divisão binária, no intestino do inseto vetor durante um período de tempo adequado de maneira que a sua diferenciação em tripomastigotas metacíclicas não aconteça precocemente (Figuras 35). Com o nocaute da TcRBP99, a expressão dos genes necessários para a proliferação dos epimastigotas é reduzida, gerando um stress nutricional que resulta na sinalização para a diferenciação em tripomastigotas metacíclicas (Figura 35). Com relação ao estudo do gene que codifica TcRBP300, os dados não permitem ainda concluir qual seria o papel dessa RBP no controle de genes envolvidos com a virulencia do parasito. Juntamente com os dados de genômica e transcriptômica

comparativa entre os clones CL Brener e CL-14 já obtidos, os estudos em andamento utilizando os clones nocautes e super-expressores de TcRBP300 poderão fornecer resultados que contribuirão para investigar esse importante aspecto da biologia do *T. cruzi.*

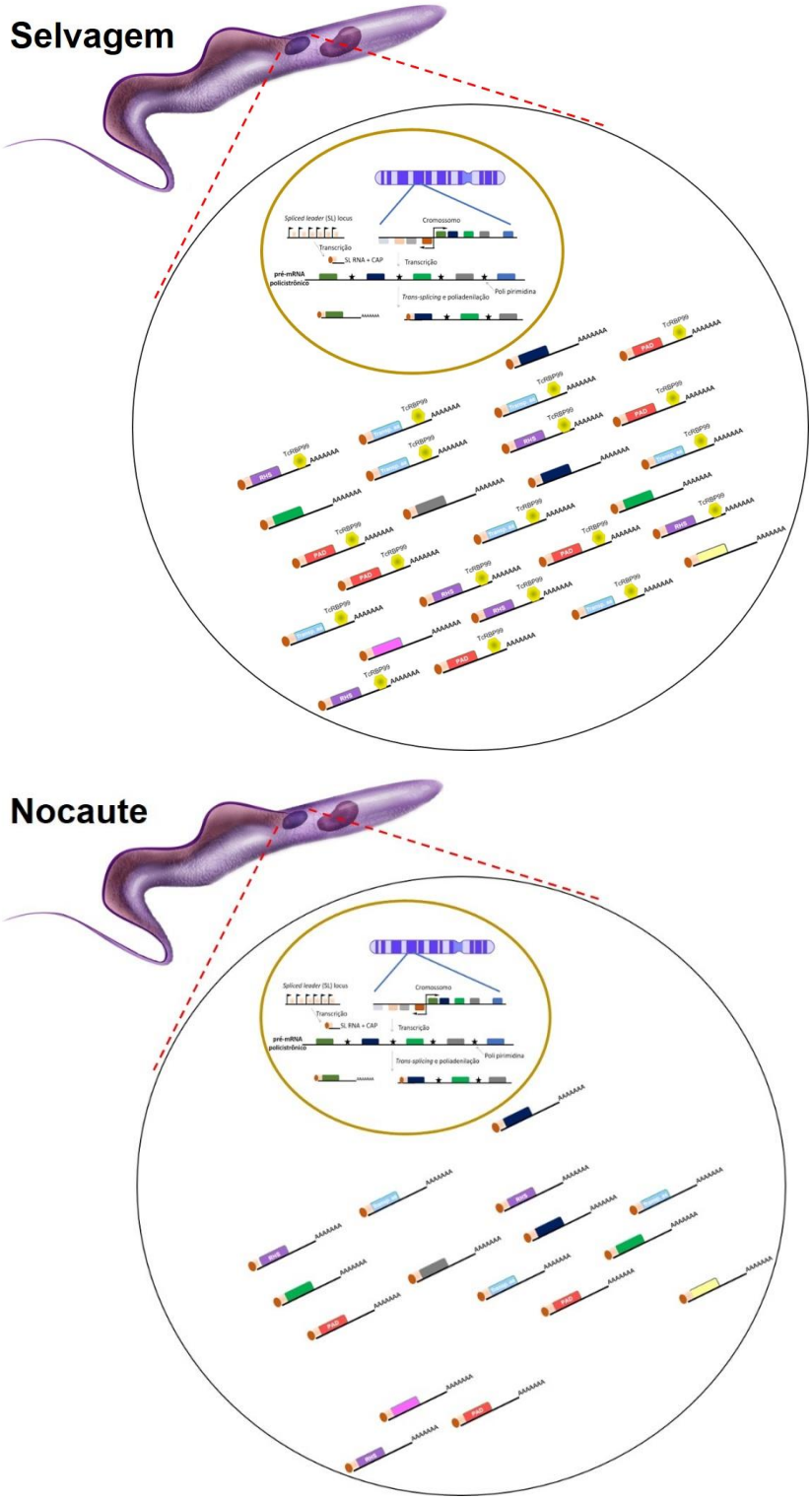


Figura 34: Representação esquemática da transcrição gênica na presença e ausência da TcRBP99. Genes do *T. cruzi* são a partir de pré-mRNAs policistrônicos que são processados através das reações de *trans-splicing* e poliadenilação gerando mRNAs maduros que são exportados para o citoplasma onde acumulam-se. No citoplasma de parasitos selvagens a proteína de ligação a RNA, TcRBP99 representada pelo hexágono amarelo, interage com mRNAs codificantes para membros da família RHS (RHS), para a proteína associada a diferenciação (PAD) e para o transportador de aminoácidos (Transp. aa), que apresentam normalmente expressão aumentada em epimastigotas, estabilizando esse mRNAs no citoplasma. Nos parasitas nocautes para a TcRBP99 baixos níveis de mRNAs codificantes para RHS, PAD e Trans. aa são encontrados no citoplasma.

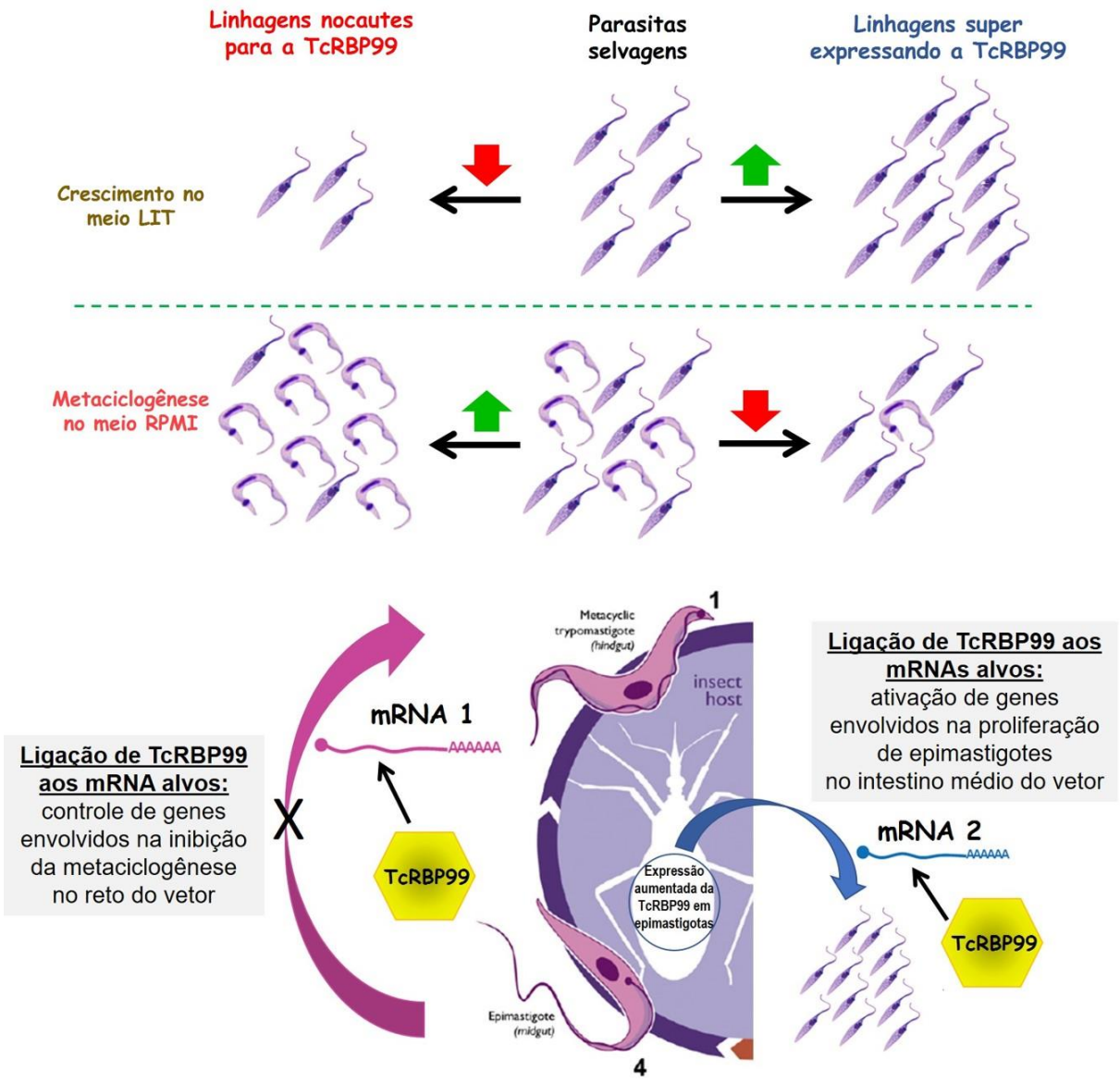


Figura 35: Esquema representando o papel da TcRBP99 em epimastigotas do *T. cruzi*.
Epimastigotas nocautes para a TcRBP99, cultivadas no meio de cultura LIT, apresentam redução do crescimento enquanto que linhagens super expressando essa RBP tem seu crescimento aumentado. Indução da diferenciação de epimastigotas em tripomastigotas metacíclicas no meio RPMI, é acelerada nos parasitos nocautes para a TcRBP99 e reduzida nos super expressores.
Em epimastigotas selvagens do clone CL Brener de *T. cruzi*, a proteína de ligação a RNA, TcRBP99, interage com mRNAs que controlam a metaciclogênese na porção final do intestino do vetor, e com mRNAs envolvidos na proliferação das epimastigotas no intestino médio do triatomíneo.

8. PERSPECTIVAS

- Realizar ensaios de infecção *in vitro* e *in vivo* com parasitas nocautes para a TcRBP99;
- Verificar se a TcRBP99 interage com gene codificante para o transportador de aminoácidos, amplificando o gene codificante para esse transportador entre os mRNAs precipitados com a proteína TcRBP99;
- Co-localizar os mRNAs diferencialmente expressos entre os as linhagens nocautes e os parasitos selvagens com a TcRBP99 por meio da técnica de FISH (Fluorescent in situ hybridization);
- Determinar a meia vida dos mRNAs diferencialmente expressos entre os as linhagens nocautes e os parasitos selvagens;
- Verificar se ocorre ao-localização da TcRBP99 em grânulos citoplasmáticos;
- Verificar a interação da TcRBP99 com regiões 3'UTR ricas em C por meio da técnica de CLIP-Seq;
- Identificar proteínas com as quais a TcRBP99 interage;
- Buscar por clones nocautes para a TcRBP300, na população selecionada com neomicina;
- Avaliar o perfil de infecção *in vitro* e *in vivo* com linhagens de CL Brener super expressando a TcRBP300 e de parasitos CL-14 nocautes para essa RBP;
- Analisar a expressão gênica global de parasitos nocautes para a TcRBP300;
- Identificar proteínas e mRNAs com os quais a TcRBP300 interage;

9. REFERENCIAS BIBLIOGRAFICAS

1. Coura JR, Vinas PA. 2010. Chagas disease: a new worldwide challenge. Nature 465:S6-7.

2. Pérez-Molina JA, Molina I. 2017. Chagas disease. The Lancet.

3. Clayton J. 2010. Chagas disease 101. Nature 465:S4-5.

4. Pereira KS, Schmidt FL, Guaraldo AM, Franco RM, Dias VL, Passos LA. 2009. Chagas' disease as a foodborne illness. J Food Prot 72:441-6.

5. Brener Z. 1973. Biology of Trypanosoma cruzi. Annu Rev Microbiol 27:347-82.

6. Cuervo P, Domont GB, De Jesus JB. 2010. Proteomics of trypanosomatids of human medical importance. J Proteomics 73:845-67.

7. Coura JR. 2015. The main sceneries of Chagas disease transmission. The vectors, blood and oral transmissions--a comprehensive review. Mem Inst Oswaldo Cruz 110:277-82.

8. Junqueira C, Caetano B, Bartholomeu DC, Melo MB, Ropert C, Rodrigues MM, Gazzinelli RT. 2010. The endless race between Trypanosoma cruzi and host immunity: lessons for and beyond Chagas disease. Expert Rev Mol Med 12:e29.

9. Kropf SP, Sa MR. 2009. The discovery of Trypanosoma cruzi and Chagas disease (1908-1909): tropical medicine in Brazil. Hist Cienc Saude Manguinhos 16 Suppl 1:13-34.

10. Souza RT, Lima FM, Barros RM, Cortez DR, Santos MF, Cordero EM, Ruiz JC, Goldenberg S, Teixeira MM, da Silveira JF. 2011. Genome size, karyotype polymorphism and chromosomal evolution in Trypanosoma cruzi. PLoS One 6:e23042.

11. Anonymous. 1999. Recommendations from a satellite meeting. Mem Inst Oswaldo Cruz 94 Suppl 1:429-32.

12. Zingales B, Souto RP, Mangia RH, Lisboa CV, Campbell DA, Coura JR, Jansen A, Fernandes O. 1998. Molecular epidemiology of American trypanosomiasis in Brazil based on dimorphisms of rRNA and mini-exon gene sequences. Int J Parasitol 28:105-12.

13. Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, Fernandes O, Guhl F, Lages-Silva E, Macedo AM, Machado CR, Miles MA, Romanha AJ, Sturm NR, Tibayrenc M, Schijman AG, Second Satellite M. 2009. A new consensus for Trypanosoma cruzi intraspecific nomenclature: second revision meeting recommends TcI to TcVI. Mem Inst Oswaldo Cruz 104:1051-4.

14. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A,

et al. 2005. The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science 309:409-15.

15. Weatherly DB, Boehlke C, Tarleton RL. 2009. Chromosome level assembly of the hybrid Trypanosoma cruzi genome. BMC Genomics 10:255.

16. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, et al. 2005. The genome of the African trypanosome Trypanosoma brucei. Science 309:416-22.

17. Toro Acevedo CA, Valente BM, Burle-Caldas GA, Galvao-Filho B, Santiago HDC, Esteves Arantes RM, Junqueira C, Gazzinelli RT, Roffe E, Teixeira SMR. 2017. Down Modulation of Host Immune Response by Amino Acid Repeats Present in a Trypanosoma cruzi Ribosomal Antigen. Front Microbiol 8:2188.

18. Franzen O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA, Andersson B. 2011. Shotgun sequencing analysis of Trypanosoma cruzi I Sylvio X10/1 and comparison with T. cruzi VI CL Brener. PLoS Negl Trop Dis 5:e984.

19. Grisard EC, Teixeira SM, de Almeida LG, Stoco PH, Gerber AL, Talavera-Lopez C, Lima OC, Andersson B, de Vasconcelos AT. 2014. Trypanosoma cruzi Clone Dm28c Draft Genome Sequence. Genome Announc 2.

20. Brener Z, Chiari E. 1963. [Morphological Variations Observed in Different Strains of Trypanosoma Cruzi]. Rev Inst Med Trop Sao Paulo 5:220-4.

21. Atayde VD, Neira I, Cortez M, Ferreira D, Freymuller E, Yoshida N. 2004. Molecular basis of non-virulence of Trypanosoma cruzi clone CL-14. Int J Parasitol 34:851-60.

22. Lima MT, Jansen AM, Rondinelli E, Gattass CR. 1991. Trypanosoma cruzi: properties of a clone isolated from CL strain. Parasitol Res 77:77-81.

23. Lima MT, Lenzi HL, Gattass CR. 1995. Negative tissue parasitism in mice injected with a noninfective clone of Trypanosoma cruzi. Parasitol Res 81:6-12.

24. Belew AT, Junqueira C, Rodrigues-Luiz GF, Valente BM, Oliveira AER, Polidoro RB, Zuccherato LW, Bartholomeu DC, Schenkman S, Gazzinelli RT, Burleigh BA, El-Sayed NM, Teixeira SMR. 2017. Comparative transcriptome profiling of virulent and non-virulent Trypanosoma cruzi underlines the role of surface proteins during infection. PLoS Pathog 13:e1006767.

25. Gunzl A, Bruderer T, Laufer G, Schimanski B, Tu LC, Chung HM, Lee PT, Lee MG. 2003. RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. Eukaryot Cell 2:542-51.

26. Teixeira SM, daRocha WD. 2003. Control of gene expression and genetic manipulation in the Trypanosomatidae. Genet Mol Res 2:148-58.

27. Vanhamme L, Pays E. 1995. Control of gene expression in trypanosomes. Microbiol Rev 59:223-40.

28. Araujo PR, Teixeira SM. 2011. Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in Trypanosoma cruzi: a review. Mem Inst Oswaldo Cruz 106:257-66.

29. LeBowitz JH, Smith HQ, Rusche L, Beverley SM. 1993. Coupling of poly(A) site selection and trans-splicing in Leishmania. Genes Dev 7:996-1007.

30. Matthews KR, Tschudi C, Ullu E. 1994. A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes Dev 8:491-501.

31. Teixeira SM, Valente BM. 2017. Mechanisms Controlling Gene Expression in Trypanosomatids. Frontiers in Parasitology: Molecular and Cellular Biology of Pathogenic Trypanosomatids 1:261-290.

32. Clayton CE. 2014. Networks of gene expression regulation in Trypanosoma brucei. Mol Biochem Parasitol 195:96-106.

33. De Gaudenzi JG, Carmona SJ, Aguero F, Frasch AC. 2013. Genome-wide analysis of 3'-untranslated regions supports the existence of post-transcriptional regulons controlling gene expression in trypanosomes. PeerJ 1:e118.

34. De Gaudenzi JG, Noe G, Campo VA, Frasch AC, Cassola A. 2011. Gene expression regulation in trypanosomatids. Essays Biochem 51:31-46.

35. Burd CG, Dreyfuss G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. Science 265:615-21.

36. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RM, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, et al. 2005. The genome of the kinetoplastid parasite, Leishmania major. Science 309:436-42.

37. De Gaudenzi J, Frasch AC, Clayton C. 2005. RNA-binding domain proteins in Kinetoplastids: a comparative analysis. Eukaryot Cell 4:2106-14.

38. Kolev NG, Ullu E, Tschudi C. 2014. The emerging role of RNA-binding proteins in the life cycle of Trypanosoma brucei. Cell Microbiol 16:482-9.

39. Kramer S, Kimblin NC, Carrington M. 2010. Genome-wide in silico screen for CCCH-type zinc finger proteins of Trypanosoma brucei, Trypanosoma cruzi and Leishmania major. BMC Genomics 11:283.

40. Oliveira C, Faoro H, Alves LR, Goldenberg S. 2017. RNA-binding proteins and their role in the regulation of gene expression in Trypanosoma cruzi and Saccharomyces cerevisiae. Genet Mol Biol 40:22-30.

41. Caro F, Bercovich N, Atorrasagasti C, Levin MJ, Vazquez MP. 2006. Trypanosoma cruzi: analysis of the complete PUF RNA-binding protein family. Exp Parasitol 113:112-24.

42. Archer SK, Luu VD, de Queiroz RA, Brems S, Clayton C. 2009. Trypanosoma brucei PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. PLoS Pathog 5:e1000565.

43. Walrad P, Paterou A, Acosta-Serrano A, Matthews KR. 2009. Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA cis-elements. PLoS Pathog 5:e1000317.

44. Dallagiovanna B, Correa A, Probst CM, Holetz F, Smircich P, de Aguiar AM, Mansur F, da Silva CV, Mortara RA, Garat B, Buck GA, Goldenberg S, Krieger MA. 2008. Functional genomic characterization of mRNAs associated with TcPUF6, a pumilio-like protein from Trypanosoma cruzi. J Biol Chem 283:8266-73.

45. Fernandez-Moya SM, Estevez AM. 2010. Posttranscriptional control and the role of RNA-binding proteins in gene regulation in trypanosomatid protozoan parasites. Wiley Interdiscip Rev RNA 1:34-46.

46. D'Orso I, Frasch AC. 2001. TcUBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes. J Biol Chem 276:34801-9.

47. D'Orso I, Frasch AC. 2002. TcUBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element- and Poly(A)-binding proteins forming a ribonucleoprotein complex. J Biol Chem 277:50520-8.

48. D'Orso I, De Gaudenzi JG, Frasch AC. 2003. RNA-binding proteins and mRNA turnover in trypanosomes. Trends Parasitol 19:151-5.

49. Kolev NG, Ramey-Butler K, Cross GA, Ullu E, Tschudi C. 2012. Developmental progression to infectivity in Trypanosoma brucei triggered by an RNA-binding protein. Science 338:1352-3.

50. Hendriks EF, Matthews KR. 2005. Disruption of the developmental programme of Trypanosoma brucei by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein. Mol Microbiol 57:706-16.

51. Hendriks EF, Robinson DR, Hinkins M, Matthews KR. 2001. A novel CCCH protein which modulates differentiation of Trypanosoma brucei to its procyclic form. EMBO J 20:6700-11.

52. Wurst M, Seliger B, Jha BA, Klein C, Queiroz R, Clayton C. 2012. Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in Trypanosoma brucei. Mol Microbiol 83:1048-63.

53. Morking PA, Rampazzo Rde C, Walrad P, Probst CM, Soares MJ, Gradia DF, Pavoni DP, Krieger MA, Matthews K, Goldenberg S, Fragoso SP, Dallagiovanna B. 2012. The zinc finger protein TcZFP2 binds target mRNAs enriched during Trypanosoma cruzi metacyclogenesis. Mem Inst Oswaldo Cruz 107:790-9.

54. Alcantara MV, Kessler RL, Goncalves REG, Marliere NP, Guarneri AA, Picchi GFA, Fragoso SP. 2018. Knockout of the CCCH zinc finger protein TcZC3H31 blocks Trypanosoma cruzi differentiation into the infective metacyclic form. Mol Biochem Parasitol 221:1-9.

55. Dallagiovanna B, Perez L, Sotelo-Silveira J, Smircich P, Duhagon MA, Garat B. 2005. Trypanosoma cruzi: molecular characterization of TcPUF6, a Pumilio protein. Exp Parasitol 109:260-4.

56. Guerra-Slompo EP, Probst CM, Pavoni DP, Goldenberg S, Krieger MA, Dallagiovanna B. 2012. Molecular characterization of the Trypanosoma cruzi specific RNA binding protein TcRBP40 and its associated mRNAs. Biochem Biophys Res Commun 420:302-7.

57. Perez-Diaz L, Correa A, Moretao MP, Goldenberg S, Dallagiovanna B, Garat B. 2012. The overexpression of the trypanosomatid-exclusive TcRBP19 RNA-binding protein affects cellular infection by Trypanosoma cruzi. Mem Inst Oswaldo Cruz 107:1076-9.

58. Perez-Diaz L, Pastro L, Smircich P, Dallagiovanna B, Garat B. 2013. Evidence for a negative feedback control mediated by the 3' untranslated region assuring the low expression level of the RNA binding protein TcRBP19 in T. cruzi epimastigotes. Biochem Biophys Res Commun 436:295-9.

59. Perez-Diaz L, Silva TC, Teixeira SM. 2017. Involvement of an RNA binding protein containing Alba domain in the stage-specific regulation of beta-amastin expression in Trypanosoma cruzi. Mol Biochem Parasitol 211:1-8.

60. Li Y, Shah-Simpson S, Okrah K, Belew AT, Choi J, Caradonna KL, Padmanabhan P, Ndegwa DM, Temanni MR, Corrada Bravo H, El-Sayed NM, Burleigh BA. 2016. Transcriptome Remodeling in Trypanosoma cruzi and Human Cells during Intracellular Infection. PLoS Pathog 12:e1005511.

61. DaRocha WD, Silva RA, Bartholomeu DC, Pires SF, Freitas JM, Macedo AM, Vazquez MP, Levin MJ, Teixeira SM. 2004. Expression of exogenous genes in Trypanosoma cruzi: improving vectors and electroporation protocols. Parasitol Res 92:113-20.

62. DaRocha WD, Otsu K, Teixeira SM, Donelson JE. 2004. Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in Trypanosoma cruzi. Mol Biochem Parasitol 133:175-86.

63. Grazielle-Silva V, Zeb TF, Bolderson J, Campos PC, Miranda JB, Alves CL, Machado CR, McCulloch R, Teixeira SM. 2015. Distinct Phenotypes Caused by Mutation of MSH2 in Trypanosome Insect and Mammalian Life Cycle Forms Are Associated with Parasite Adaptation to Oxidative Stress. PLoS Negl Trop Dis 9:e0003870.

64. Otsu K, Donelson JE, Kirchhoff LV. 1993. Interruption of a Trypanosoma cruzi gene encoding a protein containing 14-amino acid repeats by targeted insertion of the neomycin phosphotransferase gene. Mol Biochem Parasitol 57:317-30.

65. Burle-Caldas Gde A, Grazielle-Silva V, Laibida LA, DaRocha WD, Teixeira SM. 2015. Expanding the tool box for genetic manipulation of Trypanosoma cruzi. Mol Biochem Parasitol 203:25-33.

66. Clayton CE. 1999. Genetic manipulation of kinetoplastida. Parasitol Today 15:372-8.

67. Lander N, Chiurillo MA, Docampo R. 2016. Genome Editing by CRISPR/Cas9: A Game Change in the Genetic Manipulation of Protists. J Eukaryot Microbiol 63:679-90.

68. Peng D, Kurup SP, Yao PY, Minning TA, Tarleton RL. 2014. CRISPR-Cas9-mediated single-gene and gene family disruption in Trypanosoma cruzi. MBio 6:e02097-14.

69. Lander N, Li ZH, Niyogi S, Docampo R. 2015. CRISPR/Cas9-Induced Disruption of Paraflagellar Rod Protein 1 and 2 Genes in Trypanosoma cruzi Reveals Their Role in Flagellar Attachment. MBio 6:e01012.

70. Zhang WW, Lypaczewski P, Matlashewski G. 2017. Optimized CRISPR-Cas9 Genome Editing for Leishmania and Its Use To Target a Multigene Family, Induce Chromosomal Translocation, and Study DNA Break Repair Mechanisms. mSphere 2.

71. Shen B, Brown KM, Lee TD, Sibley LD. 2014. Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9. MBio 5:e01114-14.

72. Camargo EP. 1964. Growth and Differentiation in Trypanosoma Cruzi. I. Origin of Metacyclic Trypanosomes in Liquid Media. Rev Inst Med Trop Sao Paulo 6:93-100.

73. Schumann Burkard G, Jutzi P, Roditi I. 2011. Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. Mol Biochem Parasitol 175:91-4.

74. Shaw AK, Kalem MC, Zimmer SL. 2016. Mitochondrial Gene Expression Is Responsive to Starvation Stress and Developmental Transition in Trypanosoma cruzi. mSphere 1.

75. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-8.

76. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.

77. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36.

78. Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166-9.

79. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat Biotechnol 29:24-6.

80. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202-8.

81. Batista JA, Teixeira SM, Donelson JE, Kirchhoff LV, de Sa CM. 1994. Characterization of a Trypanosoma cruzi poly(A)-binding protein and its genes. Mol Biochem Parasitol 67:301-12.

82. Maris C, Dominguez C, Allain FH. 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 272:2118-31.

83. Umaer K, Williams N. 2015. Kinetoplastid Specific RNA-Protein Interactions in Trypanosoma cruzi Ribosome Biogenesis. PLoS One 10:e0131323.

84. Minning TA, Weatherly DB, Atwood J, 3rd, Orlando R, Tarleton RL. 2009. The steady-state transcriptome of the four major life-cycle stages of Trypanosoma cruzi. BMC Genomics 10:370.

85. Smircich P, Eastman G, Bispo S, Duhagon MA, Guerra-Slompo EP, Garat B, Goldenberg S, Munroe DJ, Dallagiovanna B, Holetz F, Sotelo-Silveira JR. 2015. Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in Trypanosoma cruzi. BMC Genomics 16:443.

86. Berna L, Chiribao ML, Greif G, Rodriguez M, Alvarez-Valin F, Robello C. 2017. Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in Trypanosoma cruzi. PeerJ 5:e3017.

87. Silber AM, Colli W, Ulrich H, Alves MJ, Pereira CA. 2005. Amino acid metabolic routes in Trypanosoma cruzi: possible therapeutic targets against Chagas' disease. Curr Drug Targets Infect Disord 5:53-64.

88. Dean S, Marchetti R, Kirk K, Matthews KR. 2009. A surface transporter family conveys the trypanosome differentiation signal. Nature 459:213-7.

89. Kim D, Chiurillo MA, El-Sayed N, Jones K, Santos MR, Porcile PE, Andersson B, Myler P, da Silveira JF, Ramirez JL. 2005. Telomere and subtelomere of Trypanosoma cruzi chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of T. cruzi telomeres. Gene 346:153-61.

90. Bringaud F, Biteau N, Melville SE, Hez S, El-Sayed NM, Leech V, Berriman M, Hall N, Donelson JE, Baltz T. 2002. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of Trypanosoma brucei. Eukaryot Cell 1:137-51.

91. Naguleswaran A, Gunasekera K, Schimanski B, Heller M, Hemphill A, Ochsenreiter T, Roditi I. 2015. Trypanosoma brucei RRM1 is a nuclear RNA-binding protein and modulator of chromatin structure. MBio 6:e00114.

92. Manger ID, Boothroyd JC. 1998. Identification of a nuclear protein in Trypanosoma brucei with homology to RNA-binding proteins from cis-splicing systems. Mol Biochem Parasitol 97:1-11.

93. Jensen BC, Ramasamy G, Vasconcelos EJ, Ingolia NT, Myler PJ, Parsons M. 2014. Extensive stage-regulation of translation revealed by ribosome profiling of Trypanosoma brucei. BMC Genomics 15:911.

94. Ouna BA, Stewart M, Helbig C, Clayton C. 2012. The Trypanosoma brucei CCCH zinc finger proteins ZC3H12 and ZC3H13. Mol Biochem Parasitol 183:184-8.

95. Morking PA, Dallagiovanna BM, Foti L, Garat B, Picchi GF, Umaki AC, Probst CM, Krieger MA, Goldenberg S, Fragoso SP. 2004. TcZFP1: a CCCH zinc finger protein of

Trypanosoma cruzi that binds poly-C oligoribonucleotides in vitro. Biochem Biophys Res Commun 319:169-77.

96. Beisang D, Bohjanen PR. 2012. Perspectives on the ARE as it turns 25 years old. Wiley Interdiscip Rev RNA 3:719-31.

97. Luz Ambrosio D, Lee JH, Panigrahi AK, Nguyen TN, Cicarelli RM, Gunzl A. 2009. Spliceosomal proteomics in Trypanosoma brucei reveal new RNA splicing factors. Eukaryot Cell 8:990-1000.

98. Lander N, Chiurillo MA, Storey M, Vercesi AE, Docampo R. 2016. CRISPR/Cas9-mediated endogenous C-terminal Tagging of Trypanosoma cruzi Genes Reveals the Acidocalcisome Localization of the Inositol 1,4,5-Trisphosphate Receptor. J Biol Chem 291:25505-25515.

99. Soares Medeiros LC, South L, Peng D, Bustamante JM, Wang W, Bunkofske M, Perumal N, Sanchez-Valdez F, Tarleton RL. 2017. Rapid, Selection-Free, High-Efficiency Genome Editing in Protozoan Parasites Using CRISPR-Cas9 Ribonucleoproteins. MBio 8.

100. Cassola A, De Gaudenzi JG, Frasch AC. 2007. Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes. Mol Microbiol 65:655-70.

Apêndice I: Paper publicado na revista Frontiers in Microbiology, descrevendo a proteína L7a do *T. cruzi* contendo fatores de virulência, gerados no período do mestrado.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 November 2017
doi: 10.3389/fmicb.2017.02188

Down Modulation of Host Immune Response by Amino Acid Repeats Present in a *Trypanosoma cruzi* Ribosomal Antigen

*Carlos A. Toro Acevedo[1†], Bruna M. Valente[1†], Gabriela A. Burle-Caldas[1], Bruno Galvão-Filho[1], Helton da C. Santiago[1], Rosa M. Esteves Arantes[2], Caroline Junqueira[3], Ricardo T. Gazzinelli[1,3], Ester Roffê[3] and Santuza M. R. Teixeira[1]**

[1] Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [2] Departamento de Patologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [3] Instituto de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil

OPEN ACCESS

Edited by:
Walderez Ornelas Dutra, Universidade Federal de Minas Gerais, Brazil

Reviewed by:
Marisa Mariel Fernandez, Instituto de Estudios de la Inmunidad Humoral Prof. Ricardo A. Margni (IDEHU) – CONICET-UBA, Argentina
Leonardo Freire-de-Lima, Universidade Federal do Rio de Janeiro, Brazil

***Correspondence:**
Santuza M. R. Teixeira
santuzat@ufmg.br

[†] These authors have contributed equally to this work.

Specialty section:
This article was submitted to Microbial Immunology, a section of the journal Frontiers in Microbiology

Received: 08 September 2017
Accepted: 25 October 2017
Published: 10 November 2017

Citation:
Toro Acevedo CA, Valente BM, Burle-Caldas GA, Galvão-Filho B, Santiago HdC, Esteves Arantes RM, Junqueira C, Gazzinelli RT, Roffê E and Teixeira SMR (2017) Down Modulation of Host Immune Response by Amino Acid Repeats Present in a Trypanosoma cruzi Ribosomal Antigen. Front. Microbiol. 8:2188. doi: 10.3389/fmicb.2017.02188

Several antigens from *Trypanosoma cruzi*, the causative agent of Chagas disease (CD), contain amino acid repeats identified as targets of the host immune response. Ribosomal proteins containing an Ala, Lys, Pro-rich repeat domain are among the *T. cruzi* antigens that are strongly recognized by antibodies from CD patients. Here we investigated the role of amino acid repeats present in the *T. cruzi* ribosomal protein L7a, by immunizing mice with recombinant versions of the full-length protein (TcRpL7a), as well as with truncated versions containing only the repetitive (TcRpL7aRep) or the non-repetitive domains (TcRpL7aΔRep). Mice immunized with full-length TcRpL7a produced high levels of IgG antibodies against the complete protein as well as against the repeat domain, whereas mice immunized with TcRpL7aΔRep or TcRpL7aRep produced very low levels or did not produce IgG antibodies against this antigen. Also in contrast to mice immunized with the full-length TcRpL7a, which produced high levels of IFN-γ, only low levels of IFN-γ or no IFN-γ were detected in cultures of splenocytes derived from mice immunized with truncated versions of the protein. After challenging with trypomastigotes, mice immunized with the TcRpL7a were partially protected against the infection whereas immunization with TcRpL7aΔRep did not alter parasitemia levels compared to controls. Strikingly, mice immunized with TcRpL7aRep displayed an exacerbated parasitemia compared to the other groups and 100% mortality after infection. Analyses of antibody production in mice that were immunized with TcRpL7aRep prior to infection showed a reduced humoral response to parasite antigens as well as against an heterologous antigen. *In vitro* proliferation assays with mice splenocytes incubated with different mitogens in the presence of TcRpL7aRep resulted in a drastic inhibition of B-cell proliferation and antibody production. Taken together, these results indicate that the repeat domain of TcRpL7a acts as an immunosuppressive factor that down regulates the host B-cell response against parasite antigens favoring parasite multiplication in the mammalian host.

Keywords: *Trypanosoma cruzi*, immune response, B-cell, immunomodulation, virulence factors, antigen, repeat domains

INTRODUCTION

Chagas disease (CD) or American trypanosomiasis, caused by the blood-borne parasite *Trypanosoma cruzi*, is a tropical neglected disease that currently affects about 10 million people in Latin America[1]. The infection begins with an acute phase characterized by high parasite numbers in the bloodstream, which, in spite of that, is usually asymptomatic. During the chronic phase, which usually begins 2–3 months after infection, parasitemia is undetectable, two distinct clinical forms can occur; an indeterminate, without symptoms, and a symptomatic form of CD which affects approximately 30% of patients, in which hidden parasites, mainly inside heart and digestive muscle cells can cause serious cardiac and/or digestives alterations (Prata, 2001; Teixeira et al., 2006; Junqueira et al., 2010; Rassi and de Rezende, 2012). The two drugs that are currently available for treatment of CD, Nifurtimox and Benzonidazol, thought to be effective only in acute phase, have several side effects (Castro et al., 2006). The parasite is transmitted to humans through the bite of a reduvidae insect vector known as "kissing bug." Although efforts toward the control of vector transmission have had a positive impact in several Latin America countries (Coura et al., 2014), development of a prophylactic vaccine remains a priority if we wish to prevent novel cases of CD in areas where infected vectors and mammalian reservoirs are still easily found.

Although a complex immune response known to involve immunoregulatory mechanisms that are associated with the pathology of CD has been thoroughly investigated, it is still far from being fully understood. Studies in experimental models of *T. cruzi* infection have shown the essential role of cytotoxic $CD8^+$-specific cells as well as of a strong humoral immune response against parasite antigens (Tarleton, 2007; Dumonteil, 2009; Rodrigues et al., 2012; Vasconcelos et al., 2014). Significant efforts toward the development of an effective vaccine against this parasite have been made using various antigens, such as cruzipain (Cazorla et al., 2010), amastigote surface protein (Nogueira et al., 2013), paraflagellar rod protein (Kurup and Tarleton, 2014), and different members of the *trans*-sialidase (TS) surface protein family (Hoft et al., 2007; Rosenberg et al., 2010) as well as using different immunization protocols that promote a Th1 response with stimulation of $CD8^+$ T cells (Pereira et al., 2015). Yet, none of these vaccine candidates has proven completely effective, probably due to the complex mechanisms that the parasite develops to escape from the host immune response.

Intensive efforts to search for *T. cruzi* antigens to be employed as targets for serodiagnosis as well as vaccine components resulted in the identification of a large number of proteins containing repeated amino acid sequences. *In silico* analyses based on the complete genome sequences of several pathogens and non-pathogenic microorganisms showed that the predicted proteome of intracellular protozoan parasites has a higher repetitive content than the proteome from extracellular parasites and free-living protists (Fankhauser et al., 2007; Mendes et al., 2013). Indeed, tandemly repeated amino acid sequences, which have been implicated with binding to host receptors as well as with immune-evasion mechanisms, are present in many surface proteins of intracellular protozoan parasites such as *Plasmodium* spp., *Leishmania* spp., and *T. cruzi*. Several *Leishmania* and *T. cruzi* proteins containing repeated amino acid motifs have been described as targets of B-cell immune response and a bias toward the expression of these proteins in the amastigote stage further suggests their involvement with intracellular parasitism (Goto et al., 2008, 2010). The *Leishmania* surface proteins A2 (Fernandes et al., 2014), HASP (Depledge et al., 2010), and PSA (Boceta et al., 2000), all of them containing large repeat domains, have been identified as antigens that are strongly recognized by antibodies from infected individuals. Among the *T. cruzi* surface proteins containing tandemly repeated amino acids known to be targets of the host immune response are B13 antigen (Duranti et al., 2012), TSs (Freitas et al., 2011), mucins (Giorgi and de Lederkremer, 2011), and the mucin-associated surface protein (MASP) (dos Santos et al., 2012). A sub-group of the TS protein family, which is encoded by the largest *T. cruzi* gene family, with more than 1,000 copies in the parasite genome (El-Sayed et al., 2005; Dc-Rubin and Schenkman, 2012), contains at its C-terminal region a amino acid repeat domain known as shed acute phase antigen (SAPA). Several studies suggested that the SAPA repeat is a T-cell independent antigen type I, which is a B-cell mitogen that acts as a diversion for the immune system, luring away antibodies from the TS catalytic site (Risso et al., 2007; Dc-Rubin and Schenkman, 2012).

Parasite ribosomal proteins containing amino acids repeats are also among highly immunogenic antigens identified in studies in which sera from individuals with leishmaniasis and CD were used (Ramírez et al., 2013; Longhi et al., 2014). One such antigen is the *T. cruzi* homolog of the eukaryotic ribosomal protein L7a, a component of the large subunit of the ribosome. We have previously shown that the Ala–Lys–Pro-rich repeat domain present at the N-terminus of the *T. cruzi* L7a is a target for antibodies from CD patients and that the number of repeats varies among different parasite strains (Cerqueira et al., 2008; Pais et al., 2008). Here, we investigated the role of the amino acid repeats present in *T. cruzi* L7a protein by generating recombinant versions of the complete antigen (TcRpL7a) as well as truncated versions containing only its repetitive (TcRpL7aRep) or the non-repetitive domain (TcRpL7aΔRep) and using the three antigens to immunize mice. After evaluating the immune response elicited in mice immunized which each antigen we asked which version of the TcRpL7a protein is able to induce protection against a challenge with a virulent *T. cruzi* strain. The result of immunization experiments together with the results of *in vitro* proliferation assays indicated that the repeat domain on *T. cruzi* L7a has immune regulatory properties that help the parasite to evade the specific host immune response.

MATERIALS AND METHODS

Mice, Parasites, and Sera Samples

All experiments were done with 6- to 8-week-old BALB/c female mice, purchased from the Centro de Bioterismo Central

[1]www.who.int/chagas/disease/en/

from the Universidade Federal de Minas Gerais. The design and methodology of all experiments involving mice were performed in accordance with the guidelines of COBEA (Brazilian College of Animal Experimentation), strictly followed the Brazilian law for "Procedures for the Scientific Use of Animals" (11.794/2008) and were approved by the animal-care ethics committee of the Federal University of Minas Gerais (CEUA, UFMG) under the protocol number 133/2014. The protocol to obtain human sera used in this study was approved by the Human Ethics Committee of the Federal University of Minas Gerais (COEP, UFMG) (protocol number 312/06). Blood form trypomastigotes (BFT) and cell culture trypomastigotes (TCT) of the CL Brener clone of *T. cruzi* were used. The BFT were maintained every 2 weeks by intraperitoneal (i.p.) inoculation of BALB/c mice with blood from infected mice. TCTs were maintained by infection of confluent LLC-MK2 cell line, every week, maintained at 37°C with 5% CO_2, in DMEM medium (Gibco) with 10% of fetal bovine serum (FBS). Homogenate of *T. cruzi* extracts (TcExtract) was prepared from trypomastigotes released from cell culture infection as follows. Supernatant from LLC-MK2 infected cells was harvested 7–8 days after infection and centrifugated at 2,000 × *g* for 5 min to remove culture cells. Next, the supernatant was centrifugated at 10,000 × *g* for 10 min and trypomastigotes in pellet were washed with phosphate-buffered saline (PBS) three times. Parasites were thermic shock-lysed by 10 cycles of freezing and thawing. The homogenate was suspended in sterile PBS and the protein amount was quantified by Bradford assay (BioRad).

Bioinformatics Analyses

Sequences retrieved from Tritrypdb[2] or GenBank were aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE). Predictions for B cells linear epitopes were carried out using BepiPred server 1.0 with a threshold of 0.50 (Larsen et al., 2006). For T-cell epitopes prediction we employed tools from the Immune epitope database (IEDB) which uses a combination of methods to predict best peptides for MHC II and I (Wang et al., 2008). Results are expressed in percentile scores and top 0.5% was selected as binders.

Plasmid Constructions and Production of Recombinant Proteins

The coding sequence for TcRpL7a (GenBank accession number AF316150) was amplified from *T. cruzi* genomic DNA purified from the CL Brener clone using the following primers 5′-CCGCTAGCCCCGGCAAGGAAGTGAAA-3′ and 5′-ATGTCGACCATTACGGCGGCAGCATC-3′ containing sites for *Nhe* I and *Sal* I restriction enzymes, respectively. A DNA fragment encoding the N-terminal 90 amino acids of TcRpL7a (TcRpL7aRep) was amplified using 5′-C CGCTAGCCCCGGCAAGGAAGTGAAA-3′ and 5′-GGCT CGAGCACAAAAGGTGAGATGGC-3′ primers containing sites for *Nhe* I and *Xho* I restriction enzymes, respectively. A DNA fragment encoding the C-terminal 227 amino acids for TcRpL7a (TcRpL7aΔRep) was amplified using 5′-G GGCTAGCATCTCACCTTTTGTGGCGC-3′ and 5′-ATGTC GACCATTACGGCGGCAGCATC-3′ primers containing *Nhe* I and *Sal* I sites, respectively. All three PCR amplified DNA products containing TcRpL7a, TcRpL7aRep, and TcRpL7aΔRep sequences were digested with compatible enzymes and cloned into pET21 (Novagen) expression vector to generate his-tagged proteins. Bacteria BL21 Star^{TM} (DE3) was used for transformation following the protocols provided by the manufacturer. After IPTG induction, recombinants histidine-tagged proteins were purified by Ni^{2+} chromatography with HisTrap column following the protocols provided by the column supplier (GE HealthCare). Recombinant proteins TcRpL7a and TcRpL7aΔRep were obtained as inclusion bodies and solubilized with 4 and 8 M urea, respectively, before applying to the Ni^{2+} column. After purification, all proteins were dialyzed against PBS. LPS was removed using Triton X-114 and LPS contamination in the proteins was determined by limulus amebocyte lysate (LAL) assay kit following the manufacturer protocol (Alka do Brasil).

Immunization, Challenge, and Cytokine Measurement

Three groups of 10 BALB/c female mice received prime-boost-boost immunizations by subcutaneous (s.c) injections of 10 μg of each of the three different recombinant proteins plus 18 μg of CpG B344 (Alpha DNA) and 30% (v/v) of alum rehydragel L.V. solution (Reheis), in a volume of 100 μl of saline solution, with an interval of 14 days. The fourth group of mice received injections with PBS plus CpG and alum. Nine days after the last boost, blood samples were collected from four animals in each group. The blood was centrifuged for 10 min at 2,700 × *g*, the sera collected, and stored at −20°C. To evaluate the cellular immune response four animals were euthanized 3 weeks after the last boost and the spleens were collected. Spleen mouse cells (SMCs) were cultured by harvesting spleen and smashing through 100 μm pore cell Strainer (Cell Strainer BD Falcon) and treated with ACK buffer (NH_4Cl 0.15 M, $KHCO_3$ 0.1 M e Na_2EDTA 0.1 M) for erythrocytes lysis, then washed twice with RPMI medium 1640 (Gibco), supplemented (complete RPMI medium) with 10% FBS (Gibco), 2 mM glutamine, 100 μg/ml streptomycin, and 100 U/ml penicillin. The viability of the cells was evaluated by using 0.2% trypan blue to discriminate between live and dead cells. 1×10^6 SMC/ml was incubated for 72 h at 37°C in 24-well plates with complete RPMI medium in the presence of 10 μg/μl of each recombinant protein. Cytokines (IFN-γ and IL-10) were measured in supernatants using R&D Systems Kits by enzyme-linked immunosorbent assays (ELISAs), following manufacturer instructions. The remaining animals were challenged with 5,000 BFT of the CL Brener clone. Parasitemia, determined by direct microscopy examination the number of parasites in 5 μl of blood taken from the animal tail, was monitored three times a week. Heart tissues were collected from the animals at the peak of the parasitemia for histopathological analyses and analyzed as described in the following sessions. Sera were also

[2]www.tritrypdb.org

collected from both groups of mice at the pick of parasitemia to determine antibody levels against total parasite extracts as indicated before.

Histopathological Analyses

Infected and TcRpL7aRep immunized/infected mice were sacrificed, the hearts were harvested, and fixed in 4% buffered paraformaldehyde. Fixed hearts were embedded in paraffin and sliced into 7 μm sections by microtome and stained with hematoxylin and eosin (H&E). Twenty fields from one heart tissue section per mice were analyzed by microscopy under 40× magnification. The numbers of intact amastigote nests and the percentages of inflammatory cellular infiltrate were determined using the ImageJ Software.

Western Blot and ELISA

For western blot analyses, proteins were first separated by SDS–PAGE, by using a 5% stacking gel and 12.5% resolving gel. Then, the proteins were transferred to nitrocellulose membranes, that were blocked in PBS-T (PBS plus 0.01% Tween 20) containing 5% of low fat milk for 1 h following by a overnight incubation at room temperature with primary antibodies diluted in PBS-T containing 1% of low fat milk. The membranes were washed and incubated with peroxidase-conjugated secondary antibodies (SoutherBiotech), diluted in PBS-T containing 1% of low fat milk, and submitted to chemoluminescent reaction using Luminata reagent (Millipore). For histidine tag detection, mouse anti-his monoclonal antibodies (GE Healthcare) were used in a 1:2,000 dilution. For specific antibody response, as primary antibodies were used a pool of sera from 10 chronic CD patients in a 1:250 dilution, or sera from mice acutely or chronically infected with *T. cruzi* in a 1:200 dilution. Serum from immunized mice, in a dilution of 1:200 also were used as primary antibodies. Secondary anti-human IgG antibodies (SoutherBiotech) was diluted 1:1,000 and secondary anti-mouse IgM or IgG antibodies (SoutherBiotech) were diluted 1:2,000. Sera from immunized or infected mice were analyzed by ELISA. Briefly, 96 wells microtiter plates (Nunc Immunoplates) were coated with TcRpL7a, TcRpL7aRep, and TcRpL7aΔRep recombinant proteins or with *T. cruzi* extract (diluted all of them at 10 μg/ml in 100 μl of carbonate buffer per well, pH 9.6) and incubating at 4°C overnight. Mice sera were diluted 1:100 in blocking buffer (PBS, 0.05% Tween 20, and 3% low fat milk) and incubated for 1 h at 37°C. Plates were incubated with peroxidase-conjugated goat anti-mouse IgM, IgG, IgG1, and IgG2a (SouthernBiotech) 1 h at 37°C, and peroxidase reactions were detected by incubation with 3,3′,5,5′-tetramethylbenzidine (TMB 1 mg/ml) reagent (SIGMA) at room temperature and optic density was measured spectrophotometrically at 450 nm. To determine the levels of non-specific IgM antibodies, plates were coated as described before with 100 μl of a 1/1,000 dilution of anti-mouse unlabeled α-IgM (SoutherBiotech) and 1:100 dilution of supernatants of cells cultures were added followed by the addition of peroxidase-conjugated anti-mouse α-IgM as described before.

Hypersensitivity Response against Ovalbumin

Two group of healthy BALB/c mice, five mice per group, were immunized s.c. with 100 μg of Ovalbumin (OVA; Sigma–Aldrich) diluted in 20 μl of saline buffer and emulsified with 30 μl of complete Freud's adjuvant (CFA; BD BioSciences). The OVA antigen was given twice with 10 days interval, and one group of mice was immunized with 10 μg TcRpL7aRep diluted in 100 μl of saline buffer, 24 h before each OVA immunization. Sera were collected at days 7 and 17 after the first immunization, and anti-OVA antibodies were determined by ELISA using plates coated with 1 μg of OVA antigen/well as described before. Hypersensitivity response was elicited in both group of mice by injecting 30 μl of heat aggregated OVA (HAO) in one rear footpad and saline buffer in the other. Footpad swelling was measured 24, 48, and 72 h after HAO injection and was compared with the basal footpad thickness. HAO was prepared by heating a 2% OVA solution in PBS at 70°C for 1 h. After cooling slowly at room temperature, the antigen was centrifuged at 225 × *g* for 10 min at 4°C, then was washed twice with PBS and resuspended to the original volume in PBS, as described elsewhere (Titus and Chiller, 1981).

In Vitro Stimulation and Flow Cytometry Analyses of Splenocytes

Two healthy 5- to 6-week-old BALB/c female mice were sacrificed and the spleens were harvested and processed for erythrocyte lysis as described before. Adherent cells were removed by incubating disrupted spleen tissue in 24 wells plated overnight at 37°C in complete RPMI medium. Next day, all non-adherent splenocytes were taken and the total number of viable cells was determined by Trypan blue exclusion. A total of 50×10^6 splenocytes were centrifuged at 2,000 × *g* for 5 min, and the pellet was resuspended in 1 ml of PBS with 5% of FBS. The cell suspension was mixed with 110 μl of 50 μM carboxyfluorescein succinimidyl ester (CFSE; Sigma–Aldrich) prepared in PBS and incubated for 5 min at room temperature, in the darkness. Cells were washed three times with 10 volumes of PBS with 5% FBS and resuspended in complete RPMI medium. CFSE labeled spleen cells were incubated at 37°C for 24 h, at a cell density of 10,000 cells/μl in complete RPMI medium (Control) or in the presence of 10 μg/ml of TcRpL7aRep. After this incubation period, the cells were stimulated with RPMI medium (Control), concanavalin A (ConA; Sigma–Aldrich) at 5 μg/ml, lipopolysaccharide (LPS; Sigma–Aldrich) at 10 μg/ml or by adding to plates that were previously coated with hamster anti-mouse CD3 IgG (BD BioSciences). To coat wells with α-CD3, 50 μl of a 5 ng/ml solution of anti-CD3 diluted in sterile PBS 1× was added to the wells and incubated at 4°C overnight. CFSE labeled spleen cells were also stimulated with ConA, LPS, or anti-CD3 in the presence of 10 μg/ml of TcRpL7aRep. Seventy-two or 120 h after stimulation cells were centrifuged at 2,000 × *g* for 5 min, washed with PBS with 2% FBS, and incubated for 1 h with Fc Block (BD BioSciences clone 2.4G2). Cells were washed with PBS, centrifuged at 2,000 × *g* for 5 min, and resuspended with 50 μl of different

anti-mouse monoclonal antibodies solutions, and incubated for 30 min at 25°C. The following fluorescent monoclonal antibodies were used: anti-CD19-APC-eFluor450 (eBioscience, clone 1D3), anti-CD3-Fluor450 (eBioscience, clone 17A2), anti-CD4-AlexaFluor700 (eBioscience, clone GK1.5), anti-CD8-PE-Cy5 (BD, clone 53-6.7), and anti-CD25-PE (Biolegend, clone 3C7). After staining with each fluorescent antibody, cells were washed, resuspended in PBS, and analyzed in a LSRFortessa^{TM} flow cytometer (BD BioSciences). Data were analyzed using the FlowJo V10 software.

Statistical Analysis

Data were analyzed using GraphPad Prism version 5.0 software and results presented as means ± SD. Statistical significance was performed using one-way ANOVA and non-parametric test followed by Bonferroni post-test. $P < 0.05$ was considered statistically significant.

RESULTS

L7a *T. cruzi* Ribosomal Protein Contains Amino Acid Repeats Not Found in Any Other Eukaryotic L7a

We have previously shown that ribosomal proteins containing amino acid repeats are important targets of the humoral immune response of patients with CD (DaRocha et al., 2002). Immunoscreening of a *T. cruzi* cDNA library showed that two clones encoding sequences homologous to ribosomal proteins containing similar amino acid repeats, named TcRpL7a and TcRpL19, are among the cDNA clones recognized by antibodies from CD patients (DaRocha et al., 2002). Sequence analyses of the *T. cruzi* genome database also showed that a total of six *T. cruzi* sequences homologous to eukaryotic ribosomal proteins contain amino acid repeats in their sequences (Pais et al., 2008). The alignment of the complete amino acid sequences of L7a genes present in the genomes of different *T. cruzi* strains with amino acid sequences from L7a of several other eukaryotes, including the related protozoan parasites *Trypanosoma brucei*, *Trypanosoma rangeli*, and *Leishmania* spp., showed that only *T. cruzi* L7a has a repetitive amino acid domain at its N-terminus (**Figure 1**). Besides being highly conserved in different *T. cruzi* strains, a similar Ala, Lys, Pro-rich repeated motif is also found in *T. cruzi* homologs of L7a, L19, and L23a ribosomal proteins (Cerqueira et al., 2008; Pais et al., 2008). It is noteworthy that, in spite of being a highly conserved protein, L7a sequences from members of the Trypanosomatidae family such as *T. rangeli*, *T. brucei*, *T. vivax*, *T. cruzi marinkellei*, *T. evansi*, and several *Leishmania* species do not present a repeat domain.

We have also previously shown that recombinant TcRpL7a in fusion with glutathione *S*-transferase (GST) or a synthetic peptide containing five TcRpL7a repeated motifs may be employed as targets in serological tests since they are strongly recognized by antibodies present in sera from patients with CD but not from patients with other parasitic diseases (Pais et al., 2008). To investigate the role of TcRpL7a repeats during the immune response elicited after infection by *T. cruzi*, we produced three recombinant version of the TcRpL7a protein: one containing only the 90 amino acid N-terminal region corresponding to the repeat domain (TcRpL7aRep, 10 kDa), the second containing a 223 amino acid C-terminal region, corresponding to the non-repetitive domain (TcRpL7aΔRep, 30 kDa), and a third version containing the complete, 319 amino acid, *T. cruzi* L7a antigen (TcRpL7a, 45 kDa). All recombinant antigens contain a histidine tag and were purified by affinity chromatography with Ni^{+2} columns (**Figure 2A**).

B-cell epitope prediction analysis by BepiPred software indicated the presence of eight linear epitopes in TcRpL7a sequence (**Figure 2B**, **Supplementary Figure S1A** and **Supplementary Table S1**) with scores higher than 0.5. However, the most relevant epitopes, scoring higher than 1.0, correspond to the complete repetitive region and one 15 mer sequence in the non-repetitive region of the protein (**Supplementary Figure S1A** and **Supplementary Table S1**). On the other hand, T-cell epitope prediction by IEDB found seven epitopes with percentile ranks below 0.5% for MHC class I, all of them, but one, located in the non-repetitive portion of the protein (**Supplementary Figure S1B** and **Supplementary Table S1**). The best binder, an 8 mer sequence with 0.1 percentile score, is predicted to bind strongly to two BALB/c MHC I molecules, H-2-Ld and H-2-Dd. For MHC II, only one peptide was found within rank percentile below 0.5% predicted to bind H2-IAd (BALB/c MHC II) (**Supplementary Table S1**) and this peptide is located in the non-repetitive region of the protein (**Supplementary Figure S1C**). Therefore, while the repetitive region of TcRpL7a seems to concentrate most B-cell epitopes, T-cell epitopes were mostly present at the non-repetitive region.

Using serum pools from individuals chronically infected with *T. cruzi* (**Figure 3B**) and from non-infected controls (**Figure 3C**), we showed that patients with CD produced IgG antibodies against the complete TcRpL7a as well as against the two truncated versions of the recombinant antigen, whereas sera from non-infected individuals have no IgG antibodies that recognize these proteins (**Figure 3C**). In contrast, BALB/c mice acutely infected with *T. cruzi* generated IgM antibodies that strongly recognize only the full-length recombinant protein (**Figure 3D**). As expected, non-infected mice do not produce IgM antibodies that recognize these antigens (**Figure 3E**) whereas mice acutely infected with *T. cruzi* produce IgG antibodies against the full-length protein (**Figure 3F**). All recombinant antigens tested positive in western blots with anti-his tag antibodies (**Figure 3A**).

Cellular and Humoral Immune Response in Mice Immunized with Different Domains of the TcRpL7a Antigen

To investigate the immune response against the *T. cruzi* L7a protein in a susceptible animal model of infection, we immunized BALB/c mice according to the protocol shown in **Figure 4A**, with all three distinct versions of purified recombinant TcRpL7a antigens (**Figure 2A**). Nine days after the last immunization, specific IgG antibodies present in sera from immunized animals were measured by ELISA. As shown in **Figure 4B**, immunization

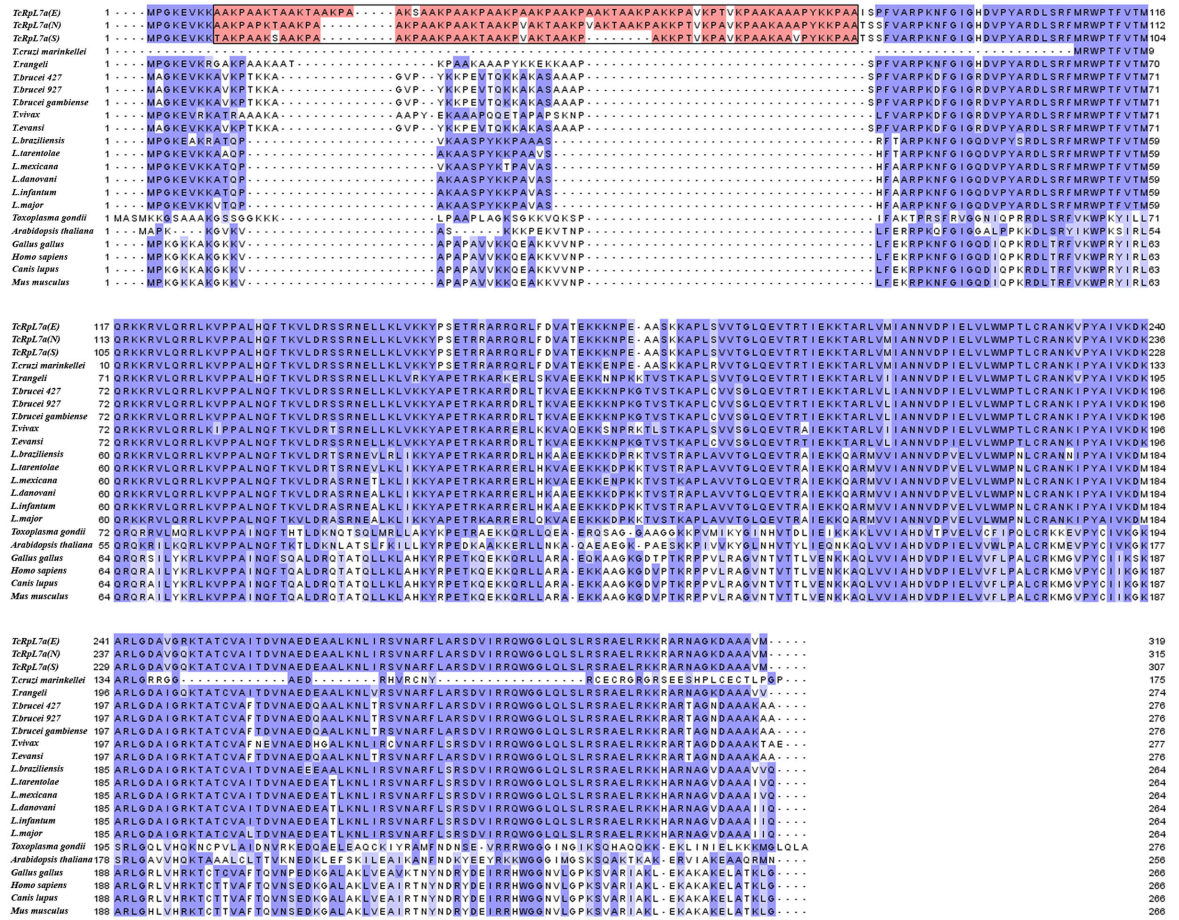


FIGURE 1 | *Trypanosoma cruzi* L7a contains amino acid repeats in its N-terminus that are absent in homologous proteins from other eukaryotes. Amino acid sequence of *T. cruzi* L7a derived from the two alleles present in the CL Brener genome [TcRpL7a(E), TcCLB.506401.320 – Esmeraldo like and TcRpL7a (N), TcCLB.510835.40 – non-Esmeraldo] and the gene present in the SylvioX-10 clone [TcRpL7a(S), TCSYLVIO_005433] were aligned and compared to L7a amino acid sequences from *T. cruzi marinkellei* (Tc_MARK_7171), *T. rangeli* (TRSC58_03561), two *T. brucei* strains (Tb427.08.1330 and Tb927.8.1330), *T. brucei gambiensis* (Tbg972.8.880), *T. vivax* (TvY486_0800730), *T. evansi* (TevSTIB805.8.1230), six *Leishmania* species (*L. brasiliensis*, LbrM.07.0560; *L. tarentolae*, LtaPcontig477-1; *L. mexicana*, LmxM.07.0500; *L. donovani*, LdBPK_070550.1; *L. infantum*, LinJ.07.0550; and *L. major*, LmjF.07.0500), *Toxoplasma gondii* (EPT25904.1) as well as from five other eukaryotes such as chicken (RL7A_CHICK P32429), dog (NM_001286939.1), mouse (NM_013721.3), plant (AL162651.1), and human (NP_000963.1). The region containing the repetitive motif AAKX, present only in *T. cruzi* L7a, is indicated by the rectangle.

with the full-length protein induces production of IgG antibodies against TcRpL7a (subtypes IgG1 and IgG2a). In contrast, mice immunized with TcRpL7aΔRep (**Figure 4C**) or with TcRpL7aRep (**Figure 4D**) did not produce significant levels of IgG antibodies against these antigens. Interestingly, immunization with the full-length version results in the production of significant levels of IgG antibodies against both TcRpL7aΔRep and TcRpL7aRep (**Figures 4E,F**). Hence, similar to the humoral immune response observed in CD patients, mice immunized with the full-length TcRpL7a antigen generate a strong humoral immune response against this protein. In contrast, mice immunized with truncated versions of the TcRpL7a antigen generated only a weak or did not elicit a detectable humoral response to the non-repetitive and the repetitive domain, respectively.

Since immunization of BALB/c mice with the truncated versions of the TcRpL7a failed to induce a humoral immune response, we investigated the cellular immune response elicited by the immunization with the different versions of the recombinant antigen. Spleen cells from mice immunized with the full-length antigen as well as with TcRpL7aΔRep and TcRpL7aRep were collected 21 days after the last immunizations and the cellular immune response was accessed by measuring IFN-γ and IL-10 in culture supernatants of splenocytes that were re-stimulated with each of the three recombinant antigens. As shown in **Figure 4G**, splenocytes from animals immunized with the full-length TcRpL7a produced high levels of IFN-γ when they were re-stimulated with the complete protein as well as with the truncated version of TcRpL7a without the repeats. Re-stimulation with the repeat domain or with the full-length antigen did not induce IFN-γ in splenocytes from animals immunized with the repeats (**Figure 4G**). However, splenocytes from animals immunized with the non-repetitive domain of TcRpL7a were able

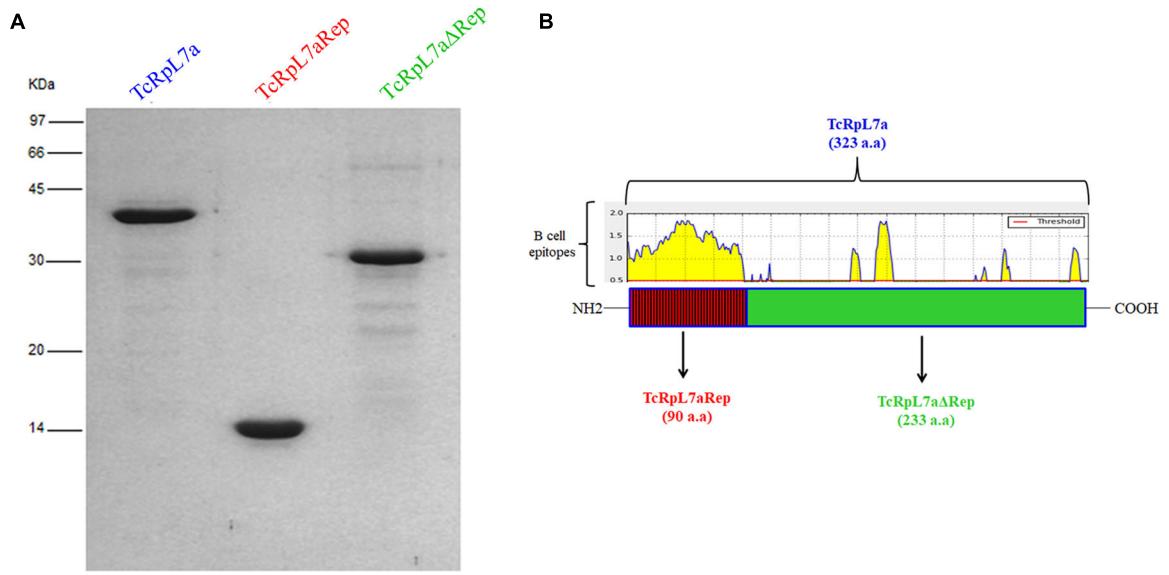


FIGURE 2 | Purified recombinant versions of TcRpL7a and B-cell epitope localization. SDS–PAGE of his-tagged recombinant full-length TcRpL7a protein and its truncated versions, TcRpL7aRep and TcRpL7aΔRep, purified by Ni-affinity chromatography **(A)**. Schematic representation of the aminoacid sequence of the TcRpL7a protein with its repetitive domain (TcRpL7aRep, in red) at the N-terminal region, and the non-repetitive domain (TcRpL7aΔRep, in green) at the C-terminal region; the localization of linear B-cell epitope prediction shown in yellow was performed with BepiPred with a threshold of 0.5 **(B)**.

to produce IFN-γ when re-stimulated with the complete antigen. Thus, in contrast to splenocytes from animals immunized with the full-length protein or with the non-repetitive domain, splenocytes from mice immunized with the repetitive domain completely failed to produce IFN-γ when re-stimulated with the truncated version of the antigen (**Figure 4G**). Similarly, only splenocytes from mice immunized with the full-length antigen (TcRpL7a) were able to produce IL-10 after being re-stimulated with the full-length antigen (**Figure 4H**). However, animals that were immunized with the truncated version containing the non-repetitive domain or with antigen containing only the repeat domain did not produce IL-10 at levels that are statistically different from the levels detected in cells from control animals that were stimulated with the different antigens (**Figure 4H**).

Immunization with Full-Length TcRpL7a Resulted in Partial Protection against a Challenge Infection Whereas Immunization with the Repeat Domain Exacerbates the Infection

Because of the strong differences observed with the immune response in mice immunized with different versions of TcRpL7a, we investigated the outcome of the infection in these animals after challenge with infective *T. cruzi* trypomastigotes. Twenty-one days after the last immunization, we challenged six animals from each group with 5,000 bloodstream trypomastigotes of the CL Brener clone and the parasitemia was monitored for 35 days, beginning 7 days after the challenge. As shown in **Figure 5A**, immunization with the full-length antigen conferred partial protection whereas immunization with TcRpL7aΔRep did not significantly alter parasitemia levels compared to control animals. A 75% reduction in parasitemia in mice immunized with the full-length antigen observed between 3 and 4 weeks after infection was corroborated by the data showing a 60% animal survival 35 days after infection compared to 30% survival of control animals. In contrast, immunization with the antigen containing only the repeated domain resulted in a twofold increase in the levels of parasitemia on day 23 compared to the control groups and eightfold increase in parasitemia compared to animals immunized with the full-length protein (**Figure 5A**). Strikingly, immunization with the repeat domain resulted in 100% mortality with a median survival of 24.5 days showing no level of protection when compared to non-vaccinated animals that showed a median survival of 27 days (**Figure 5B**). In contrast, animals vaccinated with the complete protein or the non-repetitive domain showed increased survival (**Figure 5B**).

Immunization with TcRpL7aRep Down Regulates the Humoral Immune Response against *T. cruzi*

To investigate the mechanisms responsible for the differences observed in the outcome of the infection when the group of infected naïve animals was compared with the group of animals that were immunized with TcRpL7aRep prior to infection, we analyzed the histopathology of heart tissues collected from both groups of animals at the peak of the parasitemia. As shown in **Figure 6A**, on day 20 post-infection, TcRpL7aRep immunized mice showed significantly higher parasitemia when compared to the control infected mice. Although histopathological analyses of heart tissues (**Figure 6B**) showed no significant difference in the numbers of amastigote nests (**Figure 6C**), a significant reduction

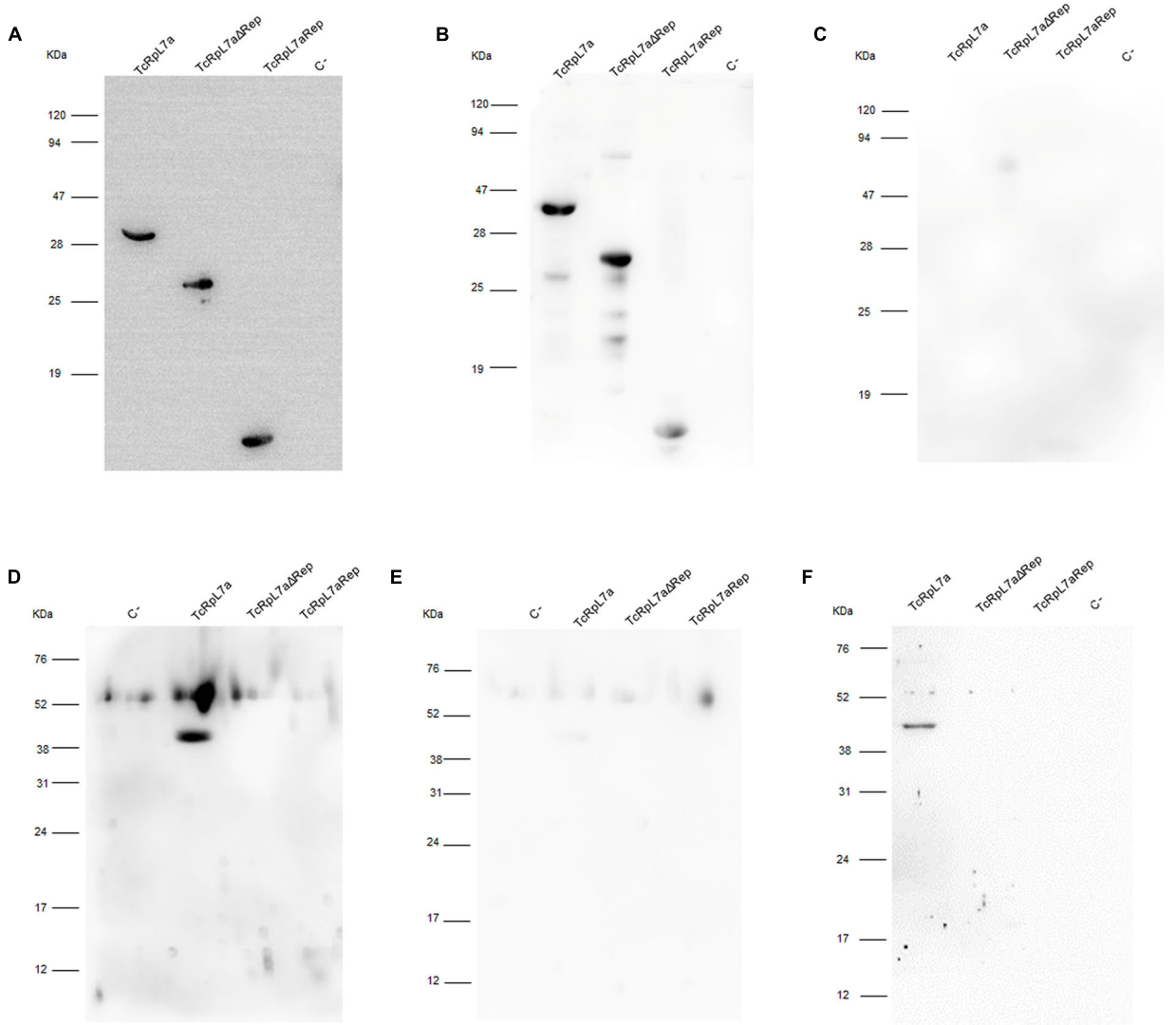


FIGURE 3 | Distinct domains of TcRpL7a are recognized by antibodies present in sera from patients with CD and from *T. cruzi*-infected mice. Western blot analyses of purified recombinant proteins TcRpL7a, TcRpL7aRep, and TcRpL7aΔRep as well as BSA, used as a negative control (C-) probed with anti-his antibody **(A)**; a pool of sera from 10 CD patients **(B)**; a pool of sera from 10 uninfected individuals **(C)**; a pool of sera from BALB/c mice acutely infected with *T. cruzi* CL Brener strain, 14 days post-infection **(D,F)**; a pool of sera from uninfected mice **(E)**. Following primary antibody incubations, membranes were incubated with horseradish peroxidase (HRP)-conjugated anti-mouse IgG **(A,F)**, HRP-conjugated anti-human IgG **(B,C)**, and HRP-conjugated anti-mouse IgM **(D,E)**. Whereas chronic CD patients have IgG antibodies that recognize all three versions of the recombinant antigen, acutelly infected mice have high levels of IgM antibodies that recognize the full-length antigen.

in the inflammatory infiltrate was observed in the hearts of mice that were previously immunized with TcRpL7aRep compared to heart tissues from infected control mice (**Figure 6D**). Also in accordance with the increased parasitemia and mortality rates of animals that were immunized with the repetitive antigen before the infection, analyses of the humoral response against total parasite antigens showed that TcRpL7aRep-immunized mice produce lower levels of total IgG and IgM anti-*T. cruzi* antibodies compared to infected control animals that were not previously immunized with this antigen (**Figures 6E,F**). It is worth mentioning that re-stimulation of splenocytes collected from both groups of mice showed no differences regarding the production of IFN-γ and IL-10 (**Supplementary Figure S2**).

Immunization with TcRpL7aRep Inhibits Humoral Response and Hypersensitivity Reaction against an Heterologous Antigen

To further investigate the immune modulatory capacity of TcRpL7aRep we analyzed the response of TcRpL7aRep-immunized mice against an heterologous antigen like OVA. Two groups of BALB/c mice were immunized with two subcutaneous

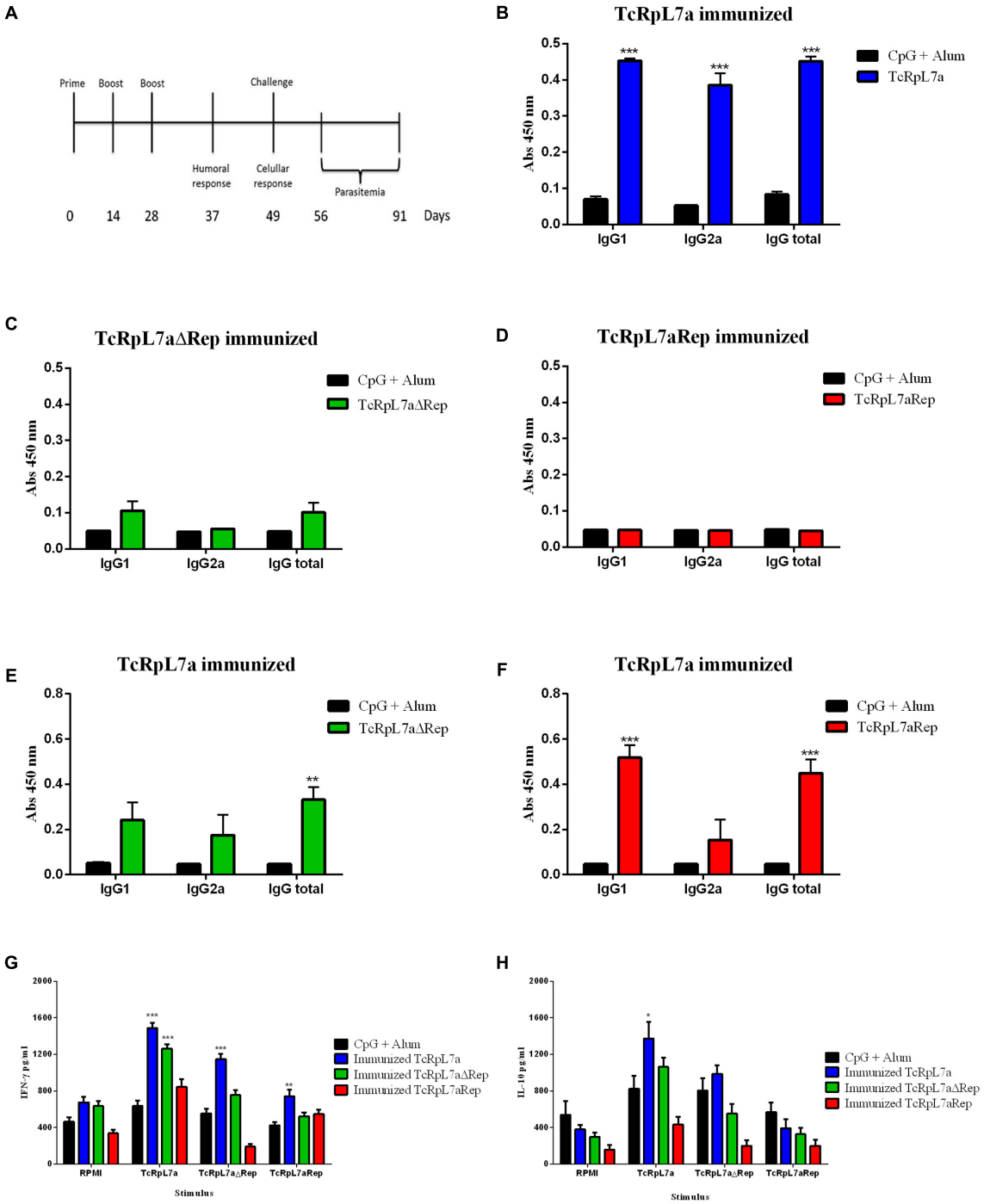


FIGURE 4 | Humoral immune response and celular activation in mice immunized with different versions of recombinantTcRpL7a. Ten BALB/c mice were immunized three times with 10 μg of recombinant proteins plus CpG and alum with an interval of 14 days. Nine days after the last immunization, sera were collected and the humoral responses evaluated. Three weeks after the last immunization, four animals in each group were sacrificed and their spleens were removed and stimulated to determine the cellular immune response. The remaining animals were challenged with 5,000 bloodstream trypomastigotes of CL Brener strain and during 35 days post-infection, parasitemia was evaluated according to the scheme shown in **(A)**. Sera levels of IgG1, IgG2a, and total IgG from mice immunized with each recombinant antigen or from control animals (CpG + Alum) were determined by ELISA using recombinant full-length TcRpL7a protein **(B)**, TcRpL7aΔRep **(C)**, or TcRpL7aRep **(D)** to coat the plates. Levels of IgG1, IgG2a, and total IgG present in sera from mice immunized with full-length TcRpL7a protein were also determined by ELISA using TcRpL7aΔRep **(E)** or TcRpL7aRep **(F)** to coat the plates. For cellular response, splenocytes from TcRpL7a-, TcRpL7aRep-, or TcRpL7aΔRep-immunized mice were incubated for 72 h in presence of one of the three versions of the antigen or in the presence of RPMI medium as a negative control. Levels of IFN-γ **(G)** and IL-10 **(H)** present in the culture supernatants were evaluated by ELISA (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$).

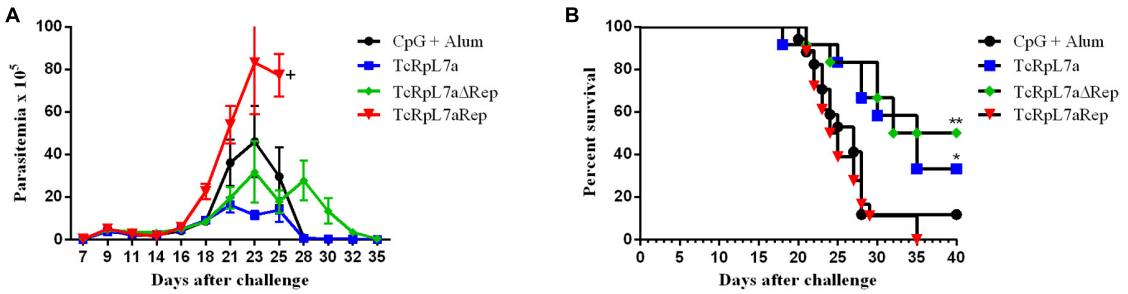


FIGURE 5 | Parasitemia and survival rates after challenging of immunized animals with *T. cruzi*. Four groups of mice were immunized with the three different versions of TcRpL7a recombinant protein or injected only with the adjuvants. Twenty-one days after the last immunization, they were challenged with 5,000 bloodstream trypomastigotes of the CL Brener strain. Parasitemia was followed for 35 days **(A)** and mortality rates **(B)** were determined in each group of mice (**$P < 0.01$; *$P < 0.05$).

injections of OVA plus CFA adjuvant. As shown in **Figure 7A**, one group has been immunized with TcRpL7aRep diluted in buffered saline, 24 h before OVA immunization, whereas the other group received saline injections before OVA immunization. Similar to the antibody response against total parasite antigens, total IgG levels directed against OVA was significantly reduced on day 7 after OVA immunization in mice that were previously injected with TcRpL7aRep compared to control mice (**Figure 7B**). Consistent with the reduction in anti-OVA IgG present in the sera of TcRpL7aRep immunized mice, these animals presented reduced footpad inflammation 24 h after injection of HAO (**Figure 7C**). Thus, besides modulating the immune response against parasite antigens, these results indicate that TcRpL7aRep can also affect the humoral and cellular response against an antigen not present in the *T. cruzi* proteome.

TcRpL7aRep Suppresses *in Vitro* B-Cell Proliferation

Hypergammaglobulinemia, a condition that characterizes the acute phase of CD, is part of the non-specific immune response that contributes to delay the specific humoral response against the parasite (Minoprio et al., 1988; Bryan et al., 2010). Several *T. cruzi* proteins such as malate dehydrogenase (Montes et al., 2002), glutamate dehydrogenase (Montes et al., 2006), rTc24 (Cordeiro da Silva et al., 1998), and proline racemase (Bryan and Norris, 2010) have been identified as polyclonal B-cell activators during *T. cruzi* infection. Members of the TS multi-gene family containing the repetitive SAPA domain at its C-terminal region have also been identified as a strong polyclonal B-cell activator that induces a strong immune response during the acute phase of CD (Gao et al., 2002). To verify whether the repeat domain present in TcRpL7a could also act as a polyclonal activator of B cells, we incubated TcRpL7aRep with mouse splenocytes *in vitro* and determined B-cell proliferation rates by analyzing CFSE dilution of CD19+ gated cells by flow cytometry. Before performing flow cytometry, we analyzed the cytotoxicity of the antigen and showed that incubation of splenocytes with 10 μg/ml of the TcRpL7aRep antigen does not significantly affect cell viability (**Supplementary Figure S3**). As shown in **Figure 8A**, in contrast to LPS, a well know polyclonal activator, incubation with TcRpL7aRep does not cause B-cell proliferation and does not induce antibody production. Interestingly, when splenocytes were incubated for 24 h with TcRpL7aRep before LPS stimulation, we observed a 60% inhibition of B-cell proliferation (**Figure 8A**) and a 50% inhibition of IgM antibody production induced by LPS (**Figure 8B**). Statistically significant differences in cell proliferation and antibody production were only observed when splenocytes were incubated with TcRpL7aRep 24 h prior the addition of LPS, suggesting that the two molecules may compete for their cell surface counterparts (**Figures 8A,B**).

To verify whether the immunosuppressive effect due to the presence of TcRpL7aRep could also be observed if B cells were activated through different mechanisms, mouse splenocytes were stimulated with α-CD3, ConA, or LPS in the presence or absence of TcRpL7aRep. As expected, α-CD3, ConA, and LPS strongly induces B-cell proliferation (**Figure 8C**). Similar to the results observed with LPS stimulation, incubation of splenocytes with TcRpL7aRep 24 h before the addition of α-CD3 and ConA totally abrogated B-cell proliferation (**Figure 8C**). Distinct from the results observed with LPS stimulation, incubation of B cells with α-CD3 and ConA resulted in a significant inhibition of proliferation if TcRpL7aRep was added at the same time of the addition of the mitogenic stimuli (**Figure 8C**). Taken together, these results indicate that TcRpL7aRep induces an anergic state of B cells, which results in a decreased capacity to proliferate in response to different stimuli or in response to parasite infection.

DISCUSSION

Trypanosoma cruzi infection elicits a complex innate and adaptive host immune response, which controls the parasite burden but often does not eradicates it. One of the biggest challenges faced by research groups studying CD is to understand the mechanisms responsible for the dynamic equilibrium established between this parasite and its host. Since there is no effective treatment for chronic infections and, despite significant advances in vector control, new infections continue to occur, the elucidation of immune evasion mechanisms elicited by

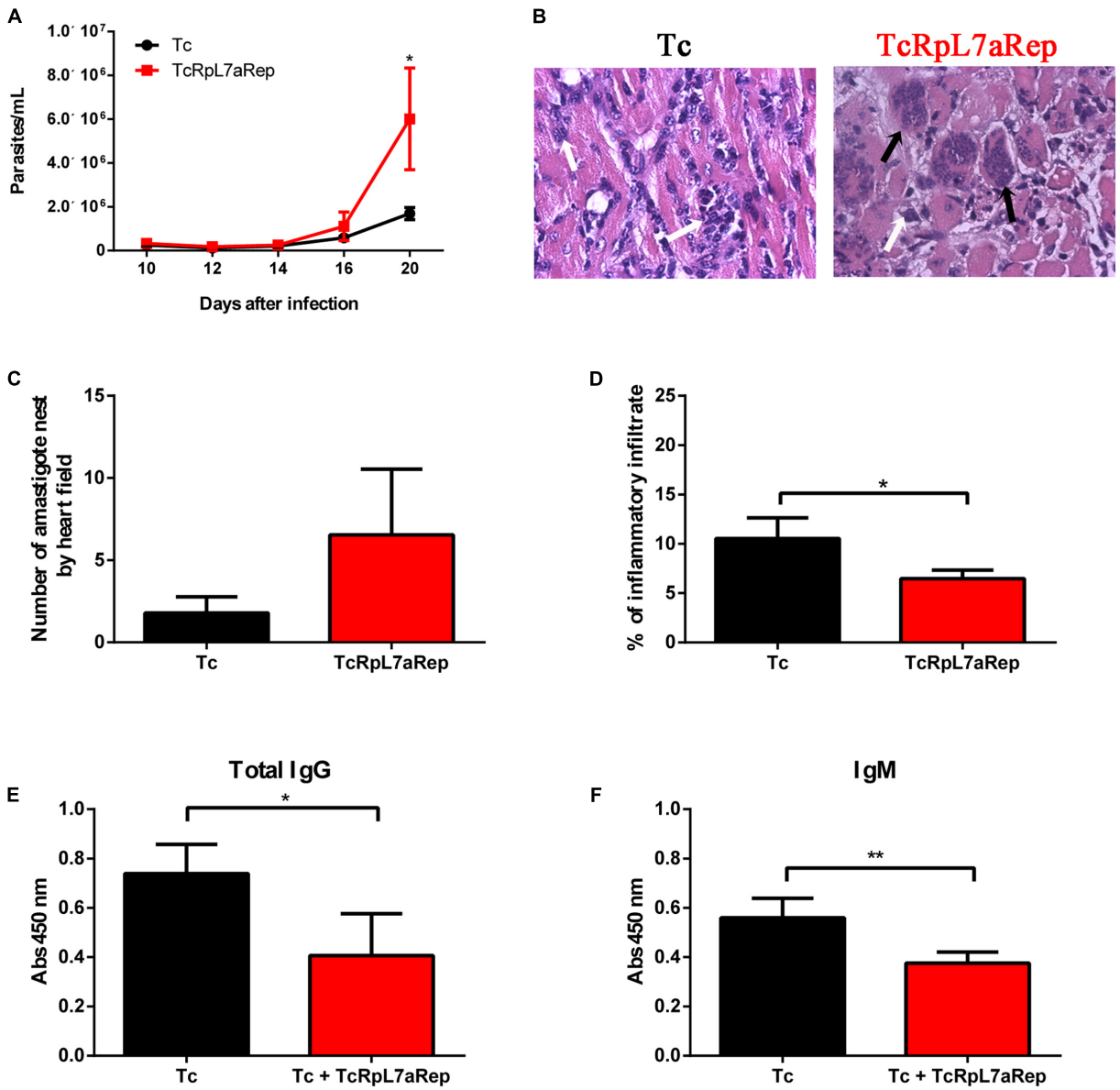


FIGURE 6 | Immunization with TcRpL7aRep down-regulates immune response in *T. cruzi*-infected mice. Two groups of BALB/c mice, immunized three times with 10 μg of recombinant TcRpL7aRep protein plus CpG and alum or only with the two adjuvants with an interval of 14 days were challenged 3 weeks after the last immunization with 5,000 bloodstream trypomastigotes of the CL Brener strain. Parasitemia of *T. cruzi*-infected mice (Tc) or TcRpL7aRep-immunized and infected mice (Tc + TcRpL7aRep) was monitored starting on day 10 post-infection **(A)**. At the peak of parasitemia (day 20), mice were euthanized, their heart and serum samples were taken for histopathological and humoral analyses, respectively. Harvested hearts were fixed on paraformaldehyde, embedded in paraffin, then sectioned, and stained with H&E. Representative images of H&E stained heart sections (40×) are shown in **(B)**, the quantification of the numbers of intact amastigote nests is shown in **(C)**, and the percentages of inflammatory infiltrate calculated by determining the percentages of area on tissue sections containing inflammatory cells are shown in **(D)**. Data are reported as the mean ± SD of 20 microscopic fields (magnification = 40×) in one H&E stained heart section per animal. White and black arrows indicate inflammatory infiltrate and integral *T. cruzi* amastigote nest, respectively. Total IgG **(E)** and IgM **(F)** antibody levels were measured by ELISA using total *T. cruzi* extract (Tc extract) to coat the plates (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$).

this parasite remains as the main prerequisite for the rational development of a vaccine against CD.

Aiming at identifying *T. cruzi* antigens that are targets of the host immune response, we and others have isolated several antigens using cDNA libraries constructed with cDNA derived from distinct parasite forms and sera from chronic CD patients (DaRocha et al., 2002; Goto et al., 2008). In these studies, it became evident that proteins containing repeat domains are immunodominant antigens that elicit a strong host humoral response. Among the *T. cruzi* antigens containing repetitive domains, we identified a group of ribosomal proteins, one of them, the homolog of the eukaryotic ribosomal protein L7a that

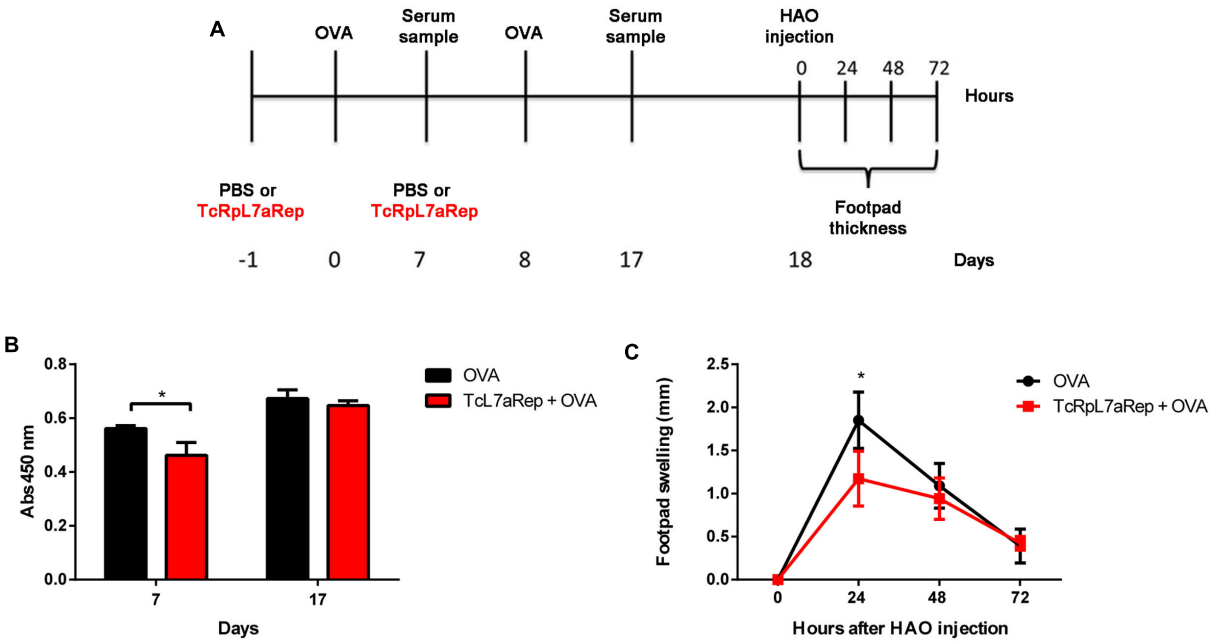


FIGURE 7 | Immunization with TcRpL7aRep impairs humoral and hypersensitivity response against the heterologous antigen OVA. Two groups of BALB/c mice were immunized subcutaneously twice with an interval of 10 days, with 100 μg of OVA diluted in 50 μl of CFA. Animals from one group of mice were injected with 10 μg of TcRpL7aRep diluted in saline buffer, 24 h before each OVA immunization whereas the control group received saline injections. Serum samples were taken at days 7 and 17, and in day 18 both groups of mice were injected with 50 μl of 2% of HAO solution in one of the footpad. Footpad swelling was measured at 24, 48, and 72 h after HAO injection, as schematized on **(A)**. Anti-OVA IgG serum levels at days 7 and 17 were measured by ELISA **(B)**. Hypersensitivity response was determined by measuring footpad swelling and subtracting the footpad thickness after HAO injection minus the basal footpad thickness **(C)** (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$).

is been used as target for sorodiagnostics. Similar studies with *Leishmania* spp., *Plasmodium falciparum*, *Neospora caninum*, and *Toxoplasma gondii*, have also shown that ribosomal proteins and proteins containing repeat domain are important targets of the immune response (Soto et al., 1995). Such strong immune response against repeat domains from various antigens lead to the "smoke screen" hypothesis, i.e., that the response against repeat domains present in parasite antigens is part of an immune-evasion mechanism that diverts the immune response away from more important parasite targets (Kemp et al., 1987). Alternatively, repetitive domains such as the SAPA domain present in the *T. cruzi* TS, which cause polyclonal activation of B cells results in depletion and/or clonal exhaustion of B cells and weakening the specific humoral host immune response (Minoprio et al., 1988; Gao et al., 2002; Bryan et al., 2010).

We have previously shown that the *T. cruzi* ribosomal protein TcRpL7a contains a repeat domain that is strongly recognized by antibodies from chronic CD patients (Pais et al., 2008). By immunizing mice with full-length protein as well as with truncated versions of the protein before challenging with *T. cruzi* trypomastigotes, we found evidence indicating that the repeat domain of TcRpL7a is part of the strategic arsenal developed by *T. cruzi* to down-modulate the host immune response and help to establish a long lived infection in the mammalian host. Similar to patients chronically infected with *T. cruzi*, mice immunized with the full-length protein produce high levels of IgG antibodies against this antigen as well as significant levels of IFN-γ and IL-10 produced by splenocytes that were re-stimulated with the full-length TcRpL7a. Mice immunized with the truncated version containing only the non-repetitive domain, where the major CD4 and CD8 epitopes are located, also produce IgG and IFN-γ, albeit at significantly low levels compared to mice immunized with TcRpL7a. Consistently, the two groups of mice that were immunized with the full-length (TcRpL7a) or with TcRpL7aΔRep were partially protected against a challenge with trypomastigotes. These results are similar to a number of studies that correlate protection levels and antibody, IFN-γ, and IL-10 production by immunized animals (Plata et al., 1984; Gazzinelli et al., 1992; Nickell et al., 1993; Junqueira et al., 2010; Roffê et al., 2012). These results are also in agreement with the presence of CD4 and CD8 epitopes in the non-repetitive domain, i.e., present in the full-length TcRpL7a and TcRpL7aΔRep recombinant proteins. In contrast, mice immunized with the truncated version containing only the repeat domain (TcRpL7aRep) have a completely different outcome, with increased parasitemia and 100% mortality compared to all other groups. It is noteworthy that although this region of the protein contains all B-cell epitopes, but lacks major CD4 or CD8 T cells epitopes, immunization with TcRpL7aRep did not result in the production of either IgG or IgM antibodies. Thus, the lack of IgM production in mice immunized with TcRpL7aRep suggests that the repeat domain has an immunomodulatory effect over B cells. Furthermore, besides having a decreased humoral response

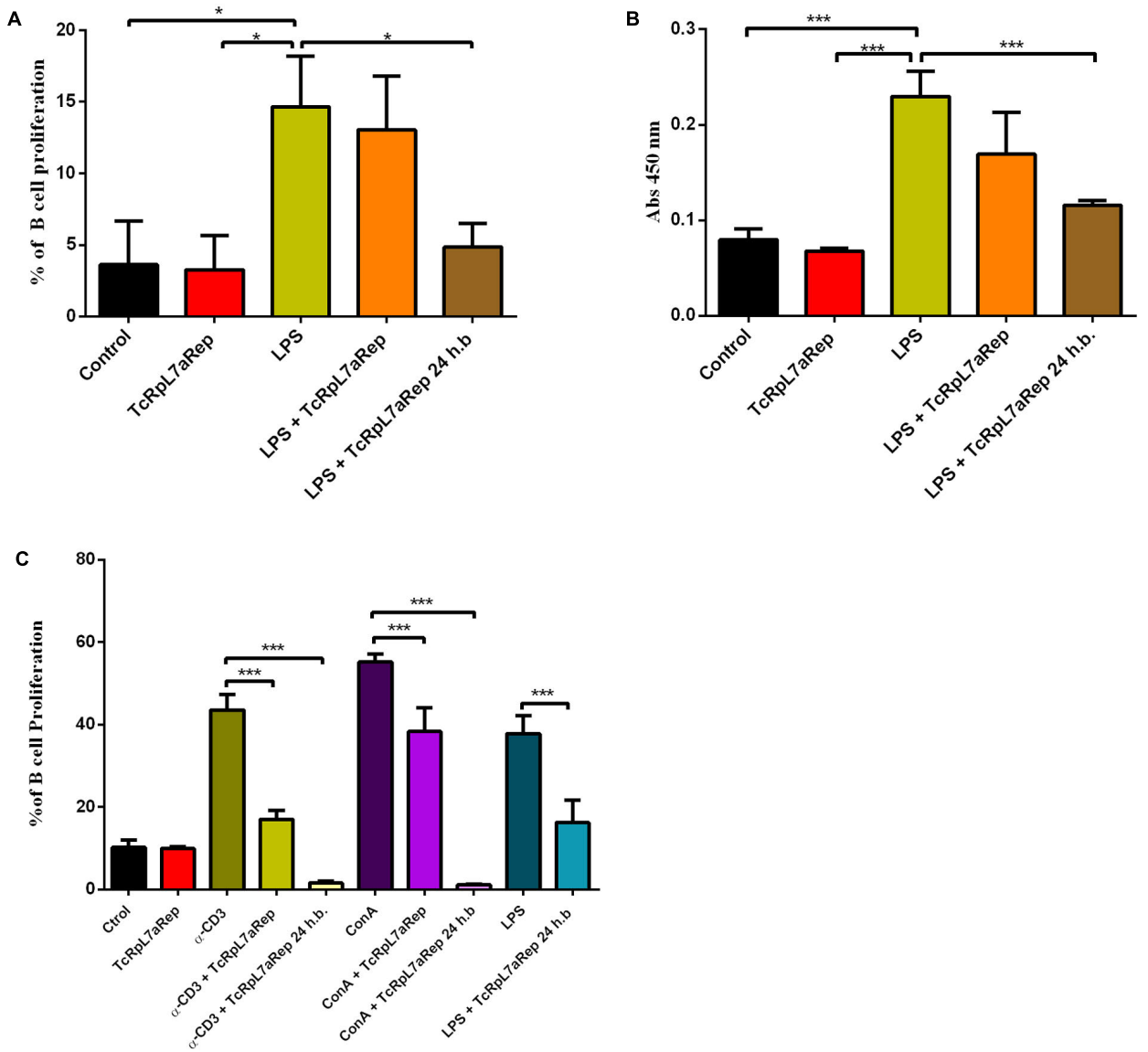


FIGURE 8 | TcRpL7aRep inhibits B-cell proliferation and specific antibody synthesis. CFSE-labeled splenocytes from healthy BALB/c mice were incubated for 72 h with TcRpL7aRep at 10 μg/ml, LPS at 10 μg/ml (control group), or with LPS in the presence of TcRpL7aRep. The recombinant antigen was added together with the stimulus or 24 h before (24 h.b.). Complete RPMI medium was used as negative control. B-cell proliferation was measured by CFSE dilution of CD19+ gated cells by flow cytometry **(A)**. Total IgM levels on cell culture supernatants were measured by ELISA **(B)**. To analyze the effect of TcRpL7aRep on B cells submitted to different stimuli, CFSE-labeled SMC from BALB/c mice were incubated for 120 h with α-CD3, ConA, or LPS in the presence or absence of TcRpL7aRep at 10 μg/ml. The repetitive antigen was added together with the stimulus or 24 h.b. B-cell proliferation was measured by flow cytometry, as indicated by the dilution of the CFSE fluorescence of gated CD19+ cells compared to positive and negative controls **(C)** (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$).

against parasite antigens, the group of mice immunized with TcRpL7aRep showed reduced inflammatory infiltrate in heart tissues even when the animals were analyzed at the pick of parasitemia.

Although the absence of CD4 and CD8 T cells epitopes in the repeat domain of TcRpL7a explains the lack of TcRpL7a-specific IFN-γ, IL-10, and antibody production in mice immunized with TcRpL7aRep, the lack of epitopes may not be the only factor responsible for the increased susceptibility to a challenging infection in animals immunized with TcRpL7aRep. It is noteworthy that cytokine production against whole parasite extract was not affected by this immunization. The results showing decreased levels of anti-*T. cruzi* IgG and IgM antibodies and a lower percentage of inflammatory infiltrate in the heart of these animals led us to hypothesize that immunization with TcRpL7aRep results in an immunosuppressive state, a hypothesis that was corroborated by the results of immunization experiments with the OVA antigen. Since immunization with

a non-related antigen also resulted in reduced humoral and hypersensitive response against the OVA antigen in animals that were previously immunized with TcRpL7aRep compared to controls, we speculate that the repeat domain of TcRpL7a acts as an immunosuppressive factor that downmodulates the host immune response in a nonspecific manner.

Based on previous studies showing that the repeat SAPA domain present in a sub-group of TSs acts as a polyclonal B-cell activator (Gao et al., 2002; Bryan et al., 2010), we performed *in vitro* proliferation assays to investigate the immunomodulatory mechanism associated with TcRpL7aRep immunization. Surprisingly, instead of acting as a polyclonal B-cell activator, the TcRpL7a repeat domain inhibited B-cell proliferation *in vitro*. As shown by *in vitro* proliferation assays of splenocytes taken from healthy mice submitted to different stimuli, incubation of these cells with TcRpL7aRep resulted in an anergic state of B cells, i.e., they became less capable to proliferate or to secrete antibodies after stimulation in response to different polyclonal activators. A possible mechanism to explain the immunosuppressive effect of TcRpL7a repeat domain is based on the immunon model, originally described by Dintzis et al. (1976). According to this model, multivalent and repetitive antigens can activate B cells without the involvement of T helper cells if the repeat domain is able to simultaneously crosslink more than 20 B-cell receptors (BCRs) (Dintzis et al., 1976, 1983). If the repeat domain contains less than 20 B-cell epitopes, which is below the threshold required for induce proliferation, a tolerance response takes place (Dintzis et al., 1983; Sulzer and Perelson, 1997). This principle was used to design drugs such as LJP-394 (abetimus sodium), comprised of four double-stranded oligodeoxyribonucleotide epitopes that induce tolerance in B cells by cross-linking surface antibodies and which has been tested as a treatment for systemic lupus erythematosus (Mosca et al., 2007). Since the repeat domain of TcRpL7a has less than 10 epitopes, we proposed that it acts as a type II T-cell independent antigen that induces B-cell tolerance. Thus, the results showing that B cells that were pre-incubated *in vitro* with TcRpL7aRep fail to proliferate or to secrete antibodies in response to α-CD3, ConA, or LPS may be due to the binding of the repeated antigen to a small number of BCR molecules. During the infection *in vivo*, this model not just explains the lack of antibody production after TcRpL7aRep immunization but also the down modulation of the immune response against parasite antigens as well as against OVA and, most importantly the increased susceptibility of infected animals that were previously immunized with TcRpL7aRep. It is likely that such significant tolerogenic effect may not result from crosslinking BCRs in one specific B-cell clone, but instead the repeat domain of TcRpL7a must be able to crosslink BCRs from several B-cell clones, or alternatively, it binds to innate like B cells, such as marginal zone B cells or B1 cells, which possess more conserved and less variable BCR (Casali and Schettino, 1996; Panda and Ding, 2015). Consistent with a hypothesis based on the tolerogen/immunon model, the difference in the protection capacity observed between the full-length TcRpL7a and the truncated version that does not carry the repeat domain may be due to the fact that immunization with the full-length antigen induces antibody production against the repeat domain TcRpL7aRep, which may prevent the binding of this region of the protein to BCRs.

How a parasite antigen that is part of the ribosome can exert this effect over host B cells is a question that has been raised since the early studies showing antibody response against conserved parasite antigens that are part of multi-component complexes (Requena et al., 2000). As part of a ribonucleoprotein complex, it can be expected that TcRpL7a might be an abundant circulating antigen that is released after parasite lysis. In light of new findings showing that microvesicles play a key role in host–parasite interaction (Ramirez et al., 2017), and that several parasite antigens including the repeat domain (N-terminal region) of MASP proteins are released by *T. cruzi* exovesicles (Lozano et al., 2017), we can also considered the possibility that TcRpL7a may be actively released from exovesicles that are shedded by bloodstream trypomastigotes. It is worth noting that proteomic analyses showed the presence of ribosomal proteins in exovesicles released by *T. cruzi* trypomastigotes (Bayer-Santos et al., 2013). It is also possible that the increased protein stability conferred by the repeat domain present in TcRpL7a, either released after parasite lysis or through exovesicles formation, contributes to sustaining the immune modulatory effect of this antigen and helps promote the immunological balance that is established between host and parasite during *T. cruzi* infection.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: CTA, BV, GB-C, ER, and ST. Performed the experiments: CTA, BV, GB-C, BG-F, CJ, REA, and ER. Analyzed the data: CTA, BV, GB-C, BG-F, CJ, REA, RG, HdCS, ST, and ER. Contributed reagents/materials/analysis tools: CJ, RG, and ST. Wrote the paper: CTA, BV, ER, HdCS, and ST.

FUNDING

This work was supported by funds from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil, no. 445456/2014-0), The Instituto Nacional de Ciencia e Tecnologia em Vacinas (INCTV), Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES-Brazil), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, no. APQ-00805-15). We are indebted to Daniella Bartholomeu for providing human serum samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02188/full#supplementary-material

FIGURE S1 | Prediction of linear B- and T-cell epitopes. Linear epitopes for B cells **(A)** and T cells MHC class I **(B)** and MHC class II **(C)** were predicted from the

amino acid sequence of TcRpL7a derived from the CL Brener strain using the software BepiPred and IEDB are underlined. The repeat domain is highlighted in bold. Scores obtained for each epitope are shown by the numbers below the lines.

FIGURE S2 | Cellular response in TcRpL7aRep-immunized *T. cruzi*-infected mice is not affected. *T. cruzi*-infected mice (Tc) or TcRpL7aRep-immunized and infected mice (Tc + TcRpL7aRep) were sacrificed at the peak of the parasitemia, and the splenocytes were cultured for 72 h in RPMI medium as negative control, or in presence of 10 μg/ml Tc extract. Levels of IFN-γ **(A)** and IL-10 **(B)** in culture supernatants were measured by ELISA (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$). No statistically significant differences in IFNg and IL-10 production by cells from infected mice that were previosuly immunized with TcRpL7a (Tc + TcRpL7aRep) compared to cells from mice that were only infected with *T. cruzi* (Tc) (IFNg, P-value = 0.8000; IL-10 P-value = 0.1000, Mann–Whitney test) were found.

FIGURE S3 | TcRpL7aRep cytotoxicity. Splenocytes from healthy BALB/c mice were incubated for 144 h with complete RPMI medium (Control) alone or in the presence of TcRpL7aRep at 10, 5, or 1 μg/ml. Cytotoxicity assays were performed by incubating cells for 2 h at 37°C with 80 μl of culture medium plus 20 μl of PBS 1× solution with 2.5 mg/ml 3-(4, 5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide (MTT) reagent. After solubilization of the formazan cristal with 10% SDS, absorbance was measured at 595 nm (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$).

TABLE S1 | T-cell and B-cell epitope predictions for the TcRpL7a antigen.

REFERENCES

Bayer-Santos, E., Aguilar-Bonavides, C., Rodrigues, S. P., Cordero, E. M., Marques, A. F., Varela-Ramirez, A., et al. (2013). Proteomic analysis of *Trypanosoma cruzi* secretome: characterization of two populations of extracellular vesicles and soluble proteins. *J. Proteome Res.* 12, 883–897. doi: 10.1021/pr300947g

Boceta, C., Alonso, C., and Jiménez-Ruiz, A. (2000). Leucine rich repeats are the main epitopes in *Leishmania infantum* PSA during canine and human visceral leishmaniasis. *Parasite Immunol.* 22, 55–62. doi: 10.1046/j.1365-3024.2000.00269.x

Bryan, M. A., Guyach, S. E., and Norris, K. A. (2010). Specific humoral immunity versus polyclonal B cell activation in *Trypanosoma cruzi* infection of susceptible and resistant mice. *PLOS Negl. Trop. Dis.* 4:e733. doi: 10.1371/journal.pntd.0000733

Bryan, M. A., and Norris, K. A. (2010). Genetic immunization converts the *Trypanosoma cruzi* B-Cell mitogen proline racemase to an effective immunogen. *Infect. Immun.* 78, 810–822. doi: 10.1128/IAI.00926-09

Casali, P., and Schettino, E. (1996). "Structure and function of natural antibodies," in *Immunology of Silicones*, eds M. Potter and N. R. Rose (Berlin: Springer), 167–179.

Castro, J. A., deMecca, M. M., and Bartel, L. C. (2006). Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). *Hum. Exp. Toxicol.* 25, 471–479. doi: 10.1191/0960327106het653oa

Cazorla, S. I., Frank, F. M., Becker, P. D., Arnaiz, M., Mirkin, G. A., Corral, R. S., et al. (2010). Redirection of the immune response to the functional catalytic domain of the cystein proteinase cruzipain improves protective immunity against *Trypanosoma cruzi* infection. *J. Infect. Dis.* 202, 136–144. doi: 10.1086/652872

Cerqueira, G. C., Bartholomeu, D. C., DaRocha, W. D., Hou, L., Freitas-Silva, D. M., Machado, C. R., et al. (2008). Sequence diversity and evolution of multigene families in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 157, 65–72. doi: 10.1016/j.molbiopara.2007.10.002

Cordeiro da Silva, A., Espinoza, A. G., Ouaissi, A., and Minoprio, P. (1998). A 24 000 MW *Trypanosoma cruzi* antigen is a B-cell activator. *Immunology* 94, 189–196. doi: 10.1046/j.1365-2567.1998.00498.x

Coura, J. R., Viñas, P. A., and Junqueira, A. C. (2014). Ecoepidemiology, short history and control of Chagas disease in the endemic countries and the new challenge for non-endemic countries. *Mem. Inst. Oswaldo Cruz* 109, 856–862. doi: 10.1590/0074-0276140236

DaRocha, W. D., Bartholomeu, D. C., Macêdo, C. D., Horta, F. M., Cunha-Neto, E., Donelson, J. E., et al. (2002). Characterization of cDNA clones encoding ribonucleoprotein antigens expressed in *Trypanosoma cruzi* amastigotes. *Parasitol. Res.* 88, 292–300. doi: 10.1007/s00436-001-0540-0

Dc-Rubin, S. S., and Schenkman, S. (2012). *Trypanosoma cruzi* trans-sialidase as a multifunctional enzyme in Chagas' disease. *Cell Microbiol.* 14, 1522–1530. doi: 10.1111/j.1462-5822.2012.01831.x

Depledge, D. P., MacLean, L. M., Hodgkinson, M. R., Smith, B. A., Jackson, A. P., Ma, S., et al. (2010). Leishmania-specific surface antigens show sub-genus sequence variation and immune recognition. *PLOS Negl. Trop. Dis.* 4:e829. doi: 10.1371/journal.pntd.0000829

Dintzis, H., Dintzis, R., and Vogelstein, B. (1976). Molecular determinants of immunogenicity: the immunon model of immune response. *Proc. Natl. Acad. Sci. U.S.A.* 73, 3671–3675. doi: 10.1073/pnas.73.10.3671

Dintzis, R., Middleton, M., and Dintzis, H. (1983). Studies on the immunogenicity and tolerogenicity of T-independent antigens. *J. Immunol.* 131, 2196–2203.

dos Santos, S. L., Freitas, L. M., Lobo, F. P., Rodrigues-Luiz, G. F., de Oliveira Mendes, T. A., Oliveira, A. C. S., et al. (2012). The MASP family of *Trypanosoma cruzi*: changes in gene expression and antigenic profile during the acute phase of experimental infection. *PLOS Negl. Trop. Dis.* 6:e1779. doi: 10.1371/journal.pntd.0001779

Dumonteil, E. (2009). Vaccine development against *Trypanosoma cruzi* and Leishmania species in the post-genomic era. *Infect. Genet. Evol.* 9, 1075–1082. doi: 10.1016/j.meegid.2009.02.009

Duranti, M., Camargo, L., Victora, G., Ianni, B., Buck, P., Mady, C., et al. (2012). Evidence for T cell help in the IgG response against tandemly repetitive *Trypanosoma cruzi* B13 protein in chronic chagas disease patients. *J. Parasitol. Res.* 2012:635873. doi: 10.1155/2012/635873

El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A.-N., et al. (2005). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309, 409–415. doi: 10.1126/science.1112631

Fankhauser, N., Nguyen-Ha, T.-M., Adler, J., and Mäser, P. (2007). Surface antigens and potential virulence factors from parasites detected by comparative genomics of perfect amino acid repeats. *Proteome sci.* 5:20. doi: 10.1186/1477-5956-5-20

Fernandes, A. P., Canavaci, A. M. C., McCall, L.-I., and Matlashewski, G. (2014). "A2 and other Visceralizing proteins of Leishmania: role in pathogenesis and application for vaccine development," in *Proteins and Proteomics of Leishmania and Trypanosoma*, eds A. Santos, M. Branquinha, C. d'Avila-Levy, L. Kneipp, and C. Sodré (Berlin: Springer), 77–101. doi: 10.1007/978-94-007-7305-9_3

Freitas, L. M., Dos Santos, S. L., Rodrigues-Luiz, G. F., Mendes, T. A., Rodrigues, T. S., Gazzinelli, R. T., et al. (2011). Genomic analyses, gene expression and antigenic profile of the trans-sialidase superfamily of *Trypanosoma cruzi* reveal an undetected level of complexity. *PLOS ONE* 6:e25914. doi: 10.1371/journal.pone.0025914

Gao, W., Wortis, H. H., and Pereira, M. A. (2002). The *Trypanosoma cruzi* trans-sialidase is a T cell-independent B cell mitogen and an inducer of non-specific Ig secretion. *Int. Immunol.* 14, 299–308. doi: 10.1093/intimm/14.3.299

Gazzinelli, R. T., Oswald, I. P., Hieny, S., James, S. L., and Sher, A. (1992). The microbicidal activity of interferon-γ-treated macrophages against *Trypanosoma cruzi* involves an l-arginine-dependent, nitrogen oxide-mediated mechanism inhibitable by interleukin-10 and transforming growth factor-β. *Eur. J. Immunol.* 22, 2501–2506. doi: 10.1002/eji.1830221006

Giorgi, M. E., and de Lederkremer, R. M. (2011). Trans-sialidase and mucins of *Trypanosoma cruzi*: an important interplay for the parasite. *Carbohydr. Res.* 346, 1389–1393. doi: 10.1016/j.carres.2011.04.006

Goto, Y., Carter, D., Guderian, J., Inoue, N., Kawazu, S.-I., and Reed, S. G. (2010). Upregulated expression of B-cell antigen family tandem repeat proteins by Leishmania amastigotes. *Infect. Immun.* 78, 2138–2145. doi: 10.1128/IAI.01102-09

Goto, Y., Carter, D., and Reed, S. G. (2008). Immunological dominance of *Trypanosoma cruzi* tandem repeat proteins. *Infect. Immun.* 76, 3967–3974. doi: 10.1128/IAI.00604-08

Hoft, D. F., Eickhoff, C. S., Giddings, O. K., Vasconcelos, J. R., and Rodrigues, M. M. (2007). Trans-sialidase recombinant protein mixed with CpG motif-containing oligodeoxynucleotide induces protective mucosal and systemic *Trypanosoma cruzi* immunity involving CD8+ CTL and B cell-mediated cross-priming. *J. Immunol.* 179, 6889–6900. doi: 10.4049/jimmunol.179.10.6889

Junqueira, C., Caetano, B., Bartholomeu, D. C., Melo, M. B., Ropert, C., Rodrigues, M. M., et al. (2010). The endless race between *Trypanosoma cruzi* and host immunity: lessons for and beyond Chagas disease. *Expert Rev. Mol. Med* 12:e29. doi: 10.1017/S1462399410001560

Kemp, D., Coppel, R., and Anders, R. (1987). Repetitive proteins and genes of malaria. *Annu. Rev. Microbiol.* 41, 181–181. doi: 10.1146/annurev.mi.41.100187.001145

Kurup, S. P., and Tarleton, R. L. (2014). The *Trypanosoma cruzi* flagellum is discarded via asymmetric cell division following invasion and provides early targets for protective CD8+ T cells. *Cell Host Microbe* 16, 439–449. doi: 10.1016/j.chom.2014.09.003

Larsen, J. E., Lund, O., and Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome Res.* 2:2. doi: 10.1186/1745-7580-2-2

Longhi, S. A., Atienza, A., Prados, G. P., Buying, A., Balouz, V., Buscaglia, C. A., et al. (2014). Cytokine production but lack of proliferation in peripheral blood mononuclear cells from chronic Chagas' disease cardiomyopathy patients in response to *T. cruzi* ribosomal P proteins. *PLOS Negl. Trop. Dis.* 8:e2906. doi: 10.1371/journal.pntd.0002906

Lozano, I. M. D., De Pablos, L. M., Longhi, S. A., Zago, M. P., Schijman, A. G., and Osuna, A. (2017). Immune complexes in chronic Chagas disease patients are formed by exovesicles from *Trypanosoma cruzi* carrying the conserved MASP N-terminal region. *Sci. Rep.* 7:44451. doi: 10.1038/srep44451

Mendes, T., Lobo, F., Rodrigues, T., Rodrigues-Luiz, G., daRocha, W., Fujiwara, R., et al. (2013). Repeat-enriched proteins are related to host cell invasion and immune evasion in parasitic protozoa. *Mol. Biol. Evol.* 30, 951–963. doi: 10.1093/molbev/mst001

Minoprio, P., Burlen, O., Pereira, P., Guilbert, B., Andrade, L., Hontebeyrie-Joskowicz, M., et al. (1988). Most B cells in acute *Trypanosoma cruzi* infection lack parasite specificity. *Scand. J. Immunol.* 28, 553–561. doi: 10.1111/j.1365-3083.1988.tb01487.x

Montes, C., Zuniga, E., Vazquez, J., Arce, C., and Gruppi, A. (2002). *Trypanosoma cruzi* mitochondrial malate dehydrogenase triggers polyclonal B-cell activation. *Clin. Exp. Immunol.* 127, 27–36. doi: 10.1046/j.1365-2249.2002.01746.x

Montes, C. L., Acosta-Rodríguez, E. V., Mucci, J., Zuniga, E. I., Campetella, O., and Gruppi, A. (2006). A *Trypanosoma cruzi* antigen signals CD11b+ cells to secrete cytokines that promote polyclonal B cell proliferation and differentiation into antibody-secreting cells. *Eur. J. Immunol.* 36, 1474–1485. doi: 10.1002/eji.200535537

Mosca, M., Baldini, C., and Bombardieri, S. (2007). LJP-394 (abetimus sodium) in the treatment of systemic lupus erythematoso. *Expert Opin. Pharmacother.* 8, 873–879. doi: 10.1517/14656566.8.6.873

Nickell, S. P., Stryker, G., and Arevalo, C. (1993). Isolation from *Trypanosoma cruzi*-infected mice of CD8+, MHC-restricted cytotoxic T cells that lyse parasite-infected target cells. *J. Immunol.* 150, 1446–1457.

Nogueira, R. T., Nogueira, A. R., Pereira, M. C. S., Rodrigues, M. M., da Costa Neves, P. C., Galler, R., et al. (2013). Recombinant yellow fever viruses elicit CD8+ T cell responses and protective immunity against *Trypanosoma cruzi*. *PLOS ONE* 8:e59347. doi: 10.1371/journal.pone.0059347

Pais, F. S., DaRocha, W. D., Almeida, R. M., Leclercq, S. Y., Penido, M. L., Fragoso, S. P., et al. (2008). Molecular characterization of ribonucleoproteic antigens containing repeated amino acid sequences from *Trypanosoma cruzi*. *Microbes Infect.* 10, 716–725. doi: 10.1016/j.micinf.2008.03.005

Panda, S., and Ding, J. L. (2015). Natural antibodies bridge innate and adaptive immunity. *J. Immunol.* 194, 13–20. doi: 10.4049/jimmunol.1400844

Pereira, I. R., Vilar-Pereira, G., Marques, V., da Silva, A. A., Caetano, B., Moreira, O. C., et al. (2015). A human type 5 adenovirus-based *Trypanosoma cruzi* therapeutic vaccine re-programs immune response and reverses chronic cardiomyopathy. *PLOS Pathog.* 11:e1004594. doi: 10.1371/journal.ppat.1004594

Plata, F., Wietzerbin, J., Pons, F. G., Falcoff, E., and Eisen, H. (1984). Synergistic protection by specific antibodies and interferon against infection by *Trypanosoma cruzi* in vitro. *Eur. J. Immunol.* 14, 930–935. doi: 10.1002/eji.1830141013

Prata, A. (2001). Clinical and epidemiological aspects of Chagas disease. *Lancet Infect. Dis.* 1, 92–100. doi: 10.1016/S1473-3099(01)00065-2

Ramírez, L., Santos, D. M., Souza, A. P., Coelho, E. A., Barral, A., Alonso, C., et al. (2013). Evaluation of immune responses and analysis of the effect of vaccination of the Leishmania major recombinant ribosomal proteins L3 or L5 in two different murine models of cutaneous leishmaniasis. *Vaccine* 31, 1312–1319. doi: 10.1016/j.vaccine.2012.12.071

Ramirez, M., Deolindo, P., Messias-Reason, I., Arigi, E. A., Choi, H., Almeida, I., et al. (2017). Dynamic flux of microvesicles modulate parasite–host cell interaction of *Trypanosoma cruzi* in eukaryotic cells. *Cell Microbiol.* 19:e12672. doi: 10.1111/cmi.12672

Rassi, A., and de Rezende, J. M. (2012). American trypanosomiasis (Chagas disease). *Infect. Dis. Clin. North Am.* 26, 275–291. doi: 10.1016/j.idc.2012.03.002

Requena, J. M. A., Alonso, C., and Soto, M. (2000). Evolutionarily conserved proteins as prominent immunogens during Leishmania infections. *Parasitol. Today* 16, 246–250. doi: 10.1016/S0169-4758(00)01651-3

Risso, M., Pitcovsky, T., Caccuri, R., Campetella, O., and Leguizamon, M. (2007). Immune system pathogenesis is prevented by the neutralization of the systemic trans-sialidase from *Trypanosoma cruzi* during severe infections. *Parasitology* 134, 503–510. doi: 10.1017/S0031182006001752

Rodrigues, M. M., Oliveira, A. C., and Bellio, M. (2012). The immune response to *Trypanosoma cruzi*: role of toll-like receptors and perspectives for vaccine development. *J. Parasitol. Res.* 2012:507874. doi: 10.1155/2012/507874

Roffê, E., Rothfuchs, A. G., Santiago, H. C., Marino, A. P. M., Ribeiro-Gomes, F. L., Eckhaus, M., et al. (2012). IL-10 limits parasite burden and protects against fatal myocarditis in a mouse model of *Trypanosoma cruzi* infection. *J. Immunol.* 188, 649–660. doi: 10.4049/jimmunol.1003845

Rosenberg, C. S., Martin, D. L., and Tarleton, R. L. (2010). CD8+ T cells specific for immunodominant trans-sialidase epitopes contribute to control of *Trypanosoma cruzi* infection but are not required for resistance. *J. Immunol.* 185, 560–568. doi: 10.4049/jimmunol.1000432

Soto, M., Requena, J., Quijada, L., Angel, S., Gomez, L., Guzman, F., et al. (1995). During active viscerocutaneous leishmaniasis the anti-P2 humoral response is specifically triggered by the parasite P proteins. *Clin. Exp. Immunol.* 100, 246–252. doi: 10.1111/j.1365-2249.1995.tb03661.x

Sulzer, B., and Perelson, A. S. (1997). Immunons revisited: binding of multivalent antigens to B cells. *Mol. Immunol.* 34, 63–74. doi: 10.1016/S0161-5890(96)00096-X

Tarleton, R. L. (2007). Immune system recognition of *Trypanosoma cruzi*. *Curr. Opin. Immunol.* 19, 430–434. doi: 10.1016/j.coi.2007.06.003

Teixeira, A. R., Nascimento, R. J., and Sturm, N. R. (2006). Evolution and pathology in Chagas disease: a review. *Memórias Instituto Oswaldo Cruz* 101, 463–491. doi: 10.1590/S0074-02762006000500001

Titus, R. G., and Chiller, J. M. (1981). A simple and effective method to assess murine delayed type hypersensitivity to proteins. *J. Immunol. Methods* 45, 65–78. doi: 10.1016/0022-1759(81)90094-6

Vasconcelos, J. R., Dominguez, M. R., Neves, R. L., Ersching, J., Araújo, A., Santos, L. I., et al. (2014). Adenovirus vector-induced CD8+ T effector memory cell differentiation and recirculation, but not proliferation, are important for protective immunity against experimental *Trypanosoma cruzi* infection. *Hum. Gene Ther.* 25, 350–363. doi: 10.1089/hum.2013.218

Wang, P., Sidney, J., Dow, C., Mothe, B., Sette, A., and Peters, B. (2008). A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLOS Comput. Biol.* 4:e1000048. doi: 10.1371/journal.pcbi.1000048

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a shared affiliation, though no other collaboration, with the authors ST, REA, RG, amd HdCS.

Apêndice II: Capítulo publicado no livro Frontiers in Parasitology, intitulado: Mechanisms Controlling Gene Expression in Trypanosomatids.

CHAPTER 6

Mechanisms Controlling Gene Expression in Trypanosomatids

Santuza M. R. Teixeira[*] and **Bruna M. Valente**

Departamento de Bioquímica e Imunologia - ICB, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Abstract: As members of a highly divergent group of eukaryotes, trypanosomatids present peculiar mechanisms of gene expression. These protozoan parasites have transcription and processing machineries that constitutively transcribe clusters of non-related genes into polycistronic pre-mRNAs, which are subsequently *trans*-spliced into monocistronic transcripts. Because of this, control of gene expression relies mainly on post-transcriptional mechanisms that are, for the most part, mediated by RNA binding proteins that control steady-state levels of mRNAs and/or their translation rates. Using primarily *Trypanosoma brucei* as a model, several groups have begun to elucidate the basic regulatory mechanisms and to define the cellular factors controlling transcription, processing, degradation and translation of mRNAs in trypanosomatids. This chapter describes studies that have been focused on a subset of genes that are differentially expressed during the life cycle of *T. brucei, T. cruzi* and few species of *Leishmania.* Although a predominance of regulatory pathways acting at a post-transcriptional level is found for most genes from all three parasites, it is also evident that the regulatory strategies chosen by different trypanosomatid species are not similar. Because of their complex and diversified gene regulatory machinery, *T. brucei, T. cruzi* and *Leishmania* spp. are able to respond rapidly to the drastic environmental changes they face during their life cycle, particularly when they move between their different hosts.

Keywords: Epigenetics, Gene expression, Polyadenylation, Post-transcription, RNA polymerase, RNA promoter, *Trans*-splicing, Transcription, *U*ntranslated region, VSG.

[*] **Address correspondence to Santuza M. R. Teixeira:** Departamento de Bioquímica e Imunologia – ICB, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; Tel: 55 31 3409 2665; E-mail: santuzat@ufmg.br

Marcelo Santos da Silva & Maria Isabel N. Cano (Eds.)

INTRODUCTION

Trypanosomatids have Unusual Mechanisms of Gene Expression

Within the *Trypanosomatidae* family, the genera *Trypanosoma* and *Leishmania* consist of several species of uniflagellated protozoa parasites that are the causative agents of several tropical diseases, such as sleeping sickness (African trypanosomiasis), caused by *Trypanosoma brucei gambiensis* and *Trypanosoma brucei rhodesiensis*; Chagas disease (American trypanosomiasis), caused by *Trypanosoma cruzi*; and different forms of leishmaniasis, caused by various species of *Leishmania*. According to recent World Health Organization surveys, it is estimated that approximately 20 million people are infected with either one of these parasites and 550 million people in the developing world are at risk of contracting leishmaniasis, Chagas disease or sleeping sickness (The World Health Report, 2014, http://www.who.ch). These digenetic parasites have distinct life cycles that present multiple differentiation forms alternating between various types of invertebrate and vertebrate hosts. Details about *Trypanosoma* and *Leishmania* life cycles and their hosts are described in Chapter 1. Therefore, the parasites must rely on multiple regulatory mechanisms to rapidly respond to the drastic environmental changes they face every time they move between their different hosts. *T. brucei, T. cruzi* and a few species of *Leishmania* are the model systems used in the studies herein described. These species belong to an ancient group of unicellular eukaryotes, which, although we tend to consider as closely related organisms, are in fact highly divergent evolutionarily [1]. Because there are still no effective drugs to treat or vaccines to prevent diseases caused by the trypanosomatids, there continues to be an urgent need for basic research that can reveal new molecular targets and parasite-specific pathways with the potential of creating new prophylactic tools and more efficient drug therapies. Details about vaccine candidates against all Tritryps and new therapeutic tools are available, respectively, on Chapters 10 and 9.

In addition to their medical relevance, trypanosomatids have been used as models for gene expression studies because of many eccentricities found in their biology. Unlike most eukaryotic genes, which are transcribed into pre-mRNAs containing exons (coding sequences) and introns (mostly non-coding sequences), trypa-

nosome protein-coding genes are intronless and are transcribed into polycistronic pre-mRNAs that are processed into mature mRNAs through "*trans*-splicing" reactions (reviewed in [2]). RNA polymerase II (RNA pol II) is the enzyme responsible for the transcription of protein-coding genes in eukaryotes. In *T. brucei*, however, two groups of protein-coding genes are transcribed by RNA polymerase I (RNA pol I), which, in most eukaryotes, exclusively transcribes ribosomal RNA genes. The co-transcriptional capping of RNAs that occurs in most eukaryotes is mediated by interactions between RNA pol II and capping enzymes. Because in trypanosomatids the addition of an m7G (cap) to the 5′ end of the primary transcript occurs during *trans*-splicing, RNA pol II-dependent transcription of protein-coding genes can be bypassed in trypanosomes. RNA pol I-dependent transcription occurs not only for a few endogenous trypanosome protein-coding genes but also for most foreign genes that can be expressed in the three groups of parasites using vectors containing RNA polymerase I promoters [3]. Indeed, except for SL genes (see below), no RNA pol II promoters have been identified in trypanosomatids. In addition to these differences regarding transcription of nuclear genes, other bizarre aspects of trypanosome biology include the unusual structure of the mitochondrial genome, called kinetoplast DNA (kDNA), and the extensive post-transcriptional modification of mitochondrial RNAs known as RNA editing, which is required to produce functional mitochondrial mRNAs [4].

Direct evidence for polycistronic transcription at a genomic level was derived from the first complete sequence analysis of chromosome 1 from *L. major*, which contains 79 genes, 29 of them encoded on one DNA strand and the remaining 50 on the opposite strand [5]. Nuclear run-on assays with strand-specific probes showed that, although a low level of nonspecific transcription likely takes place over the entire chromosome, RNA pol II-mediated transcription initiates within the strand-switch region (between the two divergent polycistronic transcription units, or PTUs) and proceeds bidirectionally towards the telomeres [6] (Fig. **1**). Sequencing of the complete *T. brucei*, *T. cruzi* and *L. major* genomes, known as the TriTryp genomes [7 - 9], revealed similar architectures in all three parasite genomes. More recent studies based on genome-wide RNA-seq analyses of the *T. brucei* genome confirmed that transcription initiation is not restricted to regions at

the beginning of gene clusters but also occurs at internal sites, albeit at lower levels [10]. However, even though a number of sequences corresponding to strand-switch regions (SSR) have been thoroughly investigated, no typical RNA pol II promoters were identified. Although the mechanism by which trypanosomes initiate transcription remains unclear, chromatin immuno-precipitation analyses indicate that transcription initiation is associated with changes in the acetylation state of histones located at the SSRs [6]. In line with these findings, genome analyses showed that, although homologs of most RNA polymerase II subunits are present in trypanosomatids, trypanosome genes encoding general transcription factors present very divergent sequences (for a review, see [11]).

In trypanosomatids, RNAP II termination must occur at a defined location between two convergent PTUs, called convergent strand-switch regions (cSSRs), to prevent the synthesis of antisense RNAs. Transcription also terminates and initiates at head-tail (HT) sites, where transcription of an upstream PTU terminates and transcription of a downstream PTU on the same strand initiates [6]. Recently, it was described that β-d-glucosyl-hydroxymethyluracil, or base J, an epigenetic modification of thymine found exclusively in Kinetoplastid [12, 13], plays a role in transcription. In *L. major*, reduction of base J resulted in genome-wide transcriptional readthrough at cSSRs and HT sites, as shown by strand-specific RT-PCR. Surprisingly, genome-wide base J regulation of RNA pol II transcription termination at cSSRs is restricted to *Leishmania* spp [14]. In *T. brucei*, base J regulates termination and gene expression at specific sites within polycistronic gene clusters, whereas, in *T. cruzi*, reduction of base J results in the formation of more active chromatin at transcription initiation sites, increased recruitment of RNA pol II and global changes in gene expression [13].

Following synthesis of polycistronic transcripts of protein-coding genes by trypanosome RNA polymerase II, processing reactions are required to generate mature mRNAs, which are monocistronic, capped and polyadenylated. Processing of mRNA through *trans*-splicing is a process that has been observed only in trypanosomatids, *Euglena* and in nematode and trematode worms [15]. Because of these differences in gene expression between trypanosomatids and higher eukaryotes, we shall first observe how the general processes of gene transcription

and mRNA maturation take place before studying how gene regulation occurs. Similarly to most eukaryotes, for efficient translation by ribosomes, the addition of a methylated G nucleotide or cap at the 5'end and the poly(A) tail at the 3'end of each mRNA is essential. However, in contrast to the general process of eukaryote transcription, in which the enzymes required for the *capping* process are associated with the C-terminal domain of the large subunit of RNA polymerase II, in trypanosomatids, transcription and mRNA processing are not interdependent events. Moreover, no consensus signal sequence for poly-adenylation has been identified in trypanosomatids, and the cap has a unique, highly methylated structure [2]. In fact, the 5'cap is part of the 39 nucleotide miniexon (or spliced leader, SL) that is joined to the 5'end of every mRNA in a transesterification reaction called *trans*-splicing. Coupling between *trans*-splicing and polyadenylation was first described by LeBowitz and collaborators [16], who demonstrated that poly(A) selection is governed by the location of the splice acceptor site of the downstream gene in the polycistronic primary transcript of the DHFR-TS locus of *L. major*. Soon after that, other groups showed that this was also the case for several *T. brucei* genes [17, 18]. Because there is no consensus polyadenylation signal found in trypanosome genes, the model proposing the existence of a conserved distance between splice acceptor site and the site for polyadenylation for the gene located upstream has become largely accepted. Several experiments using deletion mutants have tested this hypothesis and shown that, in addition to the correct distance, the presence of a polypyrimidine-rich motif is also crucial because only AG dinucleotides situated downstream from a polypyrimidine tract (PPT) are used as splice acceptor site (Fig. **1**) [19].

After the completion of the TriTryp genomes, genome-wide analyses of mRNA in *T. brucei* [10, 20, 21] and *L. major* [22] precisely identified the 5' and 3' boundaries for almost all *T. brucei* and *Leishmania* genes, confirming the rules that govern *trans*-splicing and polyadenylation. These studies also allowed for the discovery of a large number of novel coding and noncoding transcripts from previously unannotated genes and, more importantly, the discovery of an unprecedented heterogeneity of pre-mRNA-processing sites. It has been determined that in *T. brucei* the PPT can be located up to 200 nt (median is 43) upstream of the splice site and has a median length of 18 nt, with a range of 7 to

79 nt [10]. In *L. major*, PPTs range from 7 to 123 nt, with a median value of 21 nt, and are located at a median distance of 64 nt from the downstream SL addition site [22]. Using a more limited dataset, it was determined that, in *T. cruzi*, the median length of PPT is 13 nt and its median distance from an SL addition site is 18 nt [23]. These distances are in agreement with the comparative analysis of the TriTryp genomes [24], which indicates that the *T. cruzi* genome is more compact, with a smaller average size of coding regions and inter-coding regions compared with *T. brucei* and *L. major*. Importantly, these analyses also showed that the high-throughput RNA sequencing studies performed with *T. brucei* and *L. major* RNAs revealed the existence of widespread alternative *trans*-splicing and polyadenylation [21, 22]. Although Dillon and collaborators [22] did not find a correlation between alternative processing and differences in mRNA levels of genes that are differentially regulated between distinct life-cycle stages of *L. major*, a major role of alternative *trans*-splicing was revealed in the study of the *T. brucei* transcriptome, in which a number of alternative *trans*-spliced mRNAs giving rise to proteins with distinct N-terminal sequences were identified [21].

Because of polycistronic transcription and the absence of conserved signals known to be involved in transcriptional regulation, studies on gene regulation in trypanosomatids led to the concept of post-transcriptional regulons, *i.e.*, networks of gene expression consisting of different genes that are coordinately regulated at particular stages of the life cycle, during the cell cycle, or in response to external signals [25, 26]. Transcripts belonging to a regulon are targets of a group of sequence-specific RNA-binding proteins (RBPs) that recognize cis-regulatory elements in their untranslated regions and mediate the synchronized control of their fate [25, 26]. Based on this assumption, several groups have performed genome-wide analyses to identify conserved regulatory elements in the untranslated regions of *T. brucei*, *T. cruzi* and *Leishmania* genes. In parallel with the identification of cis-acting elements, *trans*-acting protein factors containing RNA-binding motifs that act as co-regulating groups of genes are beginning to be characterized. Notwithstanding, it should be mentioned that recent studies are also starting to reveal a significant role for chromatin structure and post-translation modification of histones as regulatory elements, that are also involved in parasite differentiation and adaptation to changes in the environment [27].

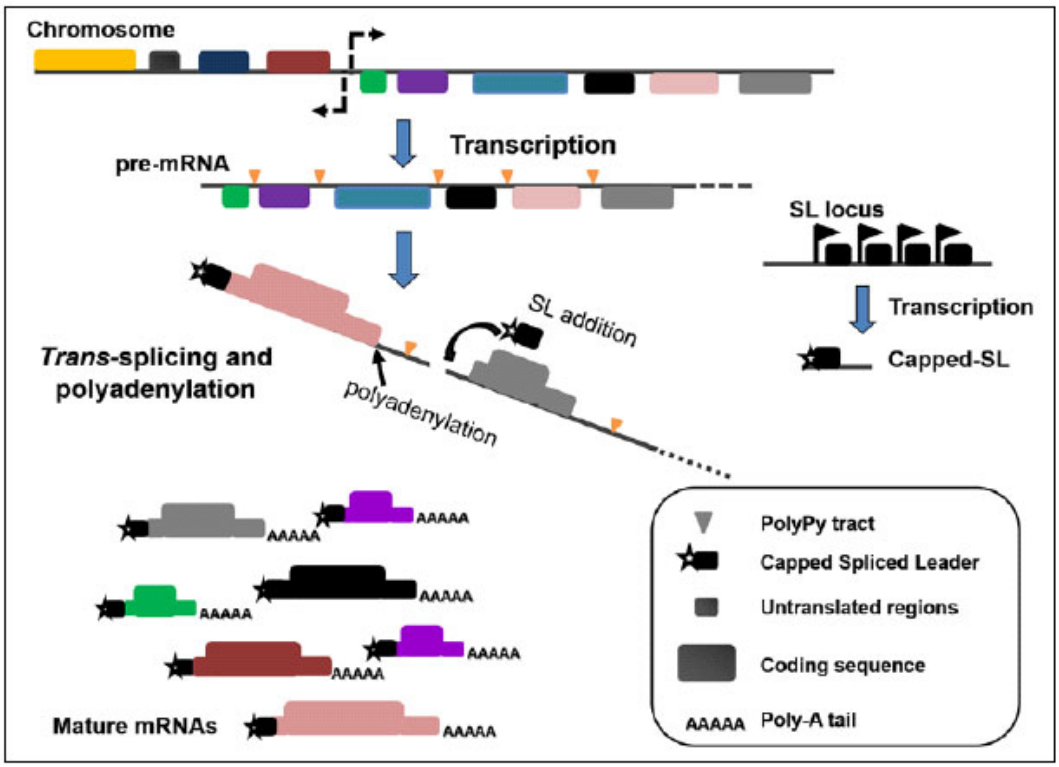


Fig. (1). The general process of mRNA synthesis in trypanosomatids. Trypanosomatid genes are organized in tandem repeats, as indicated by the boxes representing clusters of genes encoded on one DNA strand followed by a cluster of genes encoded on the opposite strand. Sequences corresponding to strand-switch regions (SSR) are represented by the two broken arrows between the two divergent gene clusters or polycistronic transcription units (PTUs). In a separate chromosomal location, there are several hundred tandem direct repeats are encoding spliced leader (SL) RNAs (small boxes), with each repeat bearing a transcriptional promoter (small flags). After transcription, the polycistronic pre-mRNA is processed by *trans*-splicing and polyadenylation reactions, which require signals present within the intergenic regions: a polypyrimidine-rich sequence (small triangles), the SL addition site and a polyadenylation addition site (arrows). Mature mRNAs composed of a coding region, untranslated regions, the capped SL sequence and the poly(A) tail are represented in the diagram at the bottom of the figure.

VSG GENES AND TRANSCRIPTIONAL CONTROL OF GENE EXPRESSION IN *T. BRUCEI*

A large number of studies aiming at revealing the molecular mechanisms involved in the control of gene expression in trypanosomatids have been performed with two groups of *T. brucei* genes encoding GPI-anchored surface proteins, variant surface glycoproteins (VSGs), which cover the bloodstream form of the parasite,

and procyclins, the major surface protein of insect forms. VSG genes are responsible for antigenic variation, perhaps one of the most interesting discoveries in the field of parasitology and a powerful survival strategy devised by African trypanosomes and few other pathogens that allows them to escape from the immunological response built by their hosts [28]. Unlike *T. cruzi* and *Leishmania*, *T. brucei* does not have an intracellular stage in its mammalian host. To survive and multiply in the mammalian bloodstream, *T. brucei* bloodstream parasites rely on the constant change of VSGs expressed on their surface. In addition to a bloodstream form (BF), *T. brucei* has another developmental stage, the procyclic form (PF), which multiplies in the insect vector, the tsetse fly. When the parasite differentiates from the bloodstream to PF in the insect midgut, the VSG coat is replaced by two insect-specific surface proteins, named GPEET and EP procyclins [29]. Thus, besides switching on and off different VSG genes during the bloodstream stage, the parasite must completely turn off VSG expression in the procyclic stage and turn on EP and GPEET procyclins in an orderly manner. Early in tsetse fly transmission, both procyclic genes are expressed, but when the trypanosomes colonize the ectoperitrophic space, GPEET is down-regulated, and EP becomes the major surface protein of late PF [29].

Initial evidence for the phenomenon of antigenic variation in *T. brucei*, described in 1910 by Ross and Thompson [30], was based on observations of peaks of parasitemia every one to two weeks in patients suffering from sleeping sickness. In 1975, the molecular basis of antigenic variation began to be uncovered through the studies by George Cross, who isolated a surface protein that was found to be the single major protein present in the BF of *T. brucei* and named it VSG for variant **s**urface glycoprotein [31]. Analyses of the N-terminal sequences of VSGs isolated from four cloned trypanosome populations derived from a single rabbit infected with *T. brucei* showed that each peak of parasitemia corresponded to a trypanosome population expressing only one VSG on their surface [32]. It soon became clear that, by switching their dense protective coat of VSG, African trypanosomes escape the host adaptive immune system. Each time a VSG switch occurs, with estimated switch rates of up to 10^{-2} cells per generation, 10^{7} identical new proteins are added to the surface of bloodstream parasites, allowing these variants to survive the host antibody response [28]. Two main questions regarding

the molecular mechanisms underlying antigenic variation are still under investigation. One important question is how trypanosomes control VSG monoallelic expression, *i.e.*, switches from the expression of VSG gene A to B while keeping all other VSG genes silent. A second question is why VSG expression occurs exclusively in BF. With the completion of the *T. brucei* genome [7], it was possible to determine that a total of 1,600 VSG genes are scattered among the 12 pairs of large chromosomes at telomeric and internal locations, as well as throughout hundreds of African-trypanosome-specific minichromosomes. The active VSG is invariably encoded in a polycistronic pre-mRNA derived from a telomeric expression site (ES) such that only one among 10-20 telomeric ES is transcribed at a time [28]. Evidence of rearrangements of VSG genes in the parasite genome was shown by early studies using restriction fragment analyses [33], and soon thereafter, it was demonstrated that the expressed VSG gene is always located near a chromosome end, approximately 1-3 kb from the repetitive telomeric hexamer sequence. Interestingly, VSG genes are not alone, instead, they are transcribed together with a gene cluster that can extend for 50-60 kb and contains approximately 10 other genes [34]. In every *T. brucei* telomere, there is a resident VSG gene. VSG genes are also found internally in chromosomes and, due to DNA rearrangement, internal VSG genes can be activated after moving to a telomeric ES location. *In situ* gene activation/inactivation is another mechanism known to assure that only one telomeric VSG is expressed in one cell. Notably, transcription of the VSG located in an ES is mediated by an RNA pol I that presents an extranucleolar location identified as a nuclear structure called an ES body (ESB) [35]. Some studies have produced a model to explain the activation/inactivation of one particular ES in a bloodstream trypanosome, comparable to silencing mechanisms described in yeast. Active VSG expression sites in *T. brucei* are known to be depleted of nucleosomes; therefore, histone-modifying enzymes must be involved [36 - 38]. Recently, a role in telomeric VSG gene silencing has been ascribed for the histone acetyltransferase (HAT3) and SIR2rp1, the only NAD-dependent histone deacetylase in *T. brucei*, as well as for other histone-modifying enzymes and chromatin chaperones, [27, 39]. Although not yet fully understood, silencing mechanisms resulting in hypoacetylated, hypomethylated compact chromatin, regulate interactions between the RNA polymerase I and VSG genes. Moreover, DNA base modification by β--

-glucosyl-hydroxymethyluracil, known as base J, also appears to be involved in telomeric gene silencing. In *T. brucei*, 50% of base J is found within telomeric repeats. Although significant levels of base J can also be found in other genomic locations in the *T. brucei* genome, its presence at silent VSG expression sites has led to the hypothesis that base J also plays a role in the control of antigenic variation. In addition, because of its absence during the tsetse life-cycle stages that do not transcribe the telomeric VSGs, base J may also contribute to shutting down all VSGs in PF [40, 41].

According to the current model, transcription of the single active VSG gene by RNA pol I, rather than RNA pol II, is a key factor that allows for monoallelic VSG expression based on its physical association with the ESB within the parasite nucleus. This unique ability of *T. brucei* to transcribe a protein-coding gene using RNA polymerase I is only possible because, as previously mentioned, trypanosome maturation of the 5'end (and the addition of cap) is independent of the RNA polymerase because, in trypanosomes, the *trans*-splicing provides the SL sequence already containing a *cap* structure at its 5'end. Moreover, because the ESB is present exclusively in bloodstream cells and has never been observed in the insect, procyclic stage, the ESB model controlling monoallelic VSG expression also provides a mechanism for repression of VSG after the bloodstream parasites differentiate into procyclic.

POST-TRANSCRIPTIONAL ELEMENTS CONTROLLING GENE EXPRESSION IN TRYPANOSOMATIDS

In the mammalian bloodstream, *T. brucei* proliferates in morphologically "slender" forms, which are able to evade host immunity by antigenic variation. Proliferating slender forms undergo morphological and molecular transformation to stumpy forms through a differentiation process that appears to be parasite density-dependent, similar to quorum-sensing systems commonly found in bacteria [42]. Once in the tsetse fly, stumpy forms, which are non-dividing and show partial adaptation to conditions in the fly midgut, differentiate into PF that are no longer infectious to mammals but can multiply in the insect vector. Thus, extensive changes in gene expression must occur when the mammalian bloodstream form differentiates into the insect PF. Recent genome-wide analyses

comparing transcript levels in *T. brucei* bloodstream and PF showed that almost 6% of all genes are life-cycle regulated by 2-fold or more [20]. However, the proteomic analysis showed that over 45% of the parasite protein levels changed between the two life cycles [43]. Procyclin genes, which encode highly abundant surface glycoproteins, are among the genes with up-regulated expression in *T. brucei* PF and that have been used as a model to study the mechanisms behind stage-specific gene expression in the parasite. Soon after bloodstream trypanosomes are taken up by a tsetse fly and differentiate into procyclic, the VSG coat is rapidly replaced by a new surface coat composed of two classes of small acidic proteins named procyclins, which are distinguished by the presence of an internal dipeptide (EP procyclins) or pentapeptide (GPEET procyclins) repeats [44]. EP and GPEET procyclins also have distinct patterns of expression, whereas early procyclic parasites transiently express GPEET procyclins, EP procyclins continue to be expressed throughout development in the insect host [44]. In addition to investigating the mechanisms behind the expression activation of GPEET and EP procyclin genes, the group led by Isabel Roditi has been investigating the factors involved in GPEET repression, which coincides with the parasite's migration to the ectoperitrophic space of the fly. Procyclin genes, initially named PARP genes, for procyclic acidic repetitive protein, are located on chromosomes VI and X and are followed by one or more procyclin-associated genes (PAGs) [45]. Similarly to VSG, procyclins are transcribed by RNA polymerase I, but, unexpectedly, procyclin and VSG promoters bear little resemblance to each other or to the classical RNA polymerase I rRNA promoter [46, 47].

Stage-specific control of procyclin gene expression occurs at several levels, including at the transcription initiation as well as transcription elongation rates of procyclic genes, *trans*-splicing efficiency, mRNA degradation and translation. Although procyclin mRNAs are not detectable in bloodstream trypanosomes, they are transcribed at low levels in bloodstream parasites, indicating that post-transcription mechanisms occur to prevent the accumulation of procyclin mRNA in BF [48]. Moreover, although procyclin and PAG genes are derived from the same polycistronic transcription unit, there is a 100-fold difference in the steady-state levels when procyclin and PAG mRNAs are compared [46]. Analysis of the

sequence of events that occurs when early PF differentiate to late PF revealed several mechanisms that could explain the differences in the accumulation of EP and GPEET procyclin mRNAs, including the role of regulatory sequences in the 3' untranslated regions (UTRs) in determining the steady-state levels of these mRNAs. In a recent study, procyclin expression during synchronous differentiation of BF - to PF, triggered *in vitro* by the addition of cis-aconitate and by dropping the temperature, was carefully analyzed by measuring steady-state levels and mRNA half-lives as well as by mapping the 3' ends of the different mRNAs. It was shown that GPEET mRNA has a half-life of more than 6 h in early differentiating PF, which is markedly reduced to 3.8 h in late PF. In contrast, EP procyclin mRNA has a half-life of 2.3 h in early PF that is increased to 6 h in late PFs [49]. These studies also indicate that differential processing of GPEET transcripts may also contribute to their repression by preventing the generation of mature mRNA in late procyclic. To identify cis-acting regulatory elements, sequences derived from the 3' UTR has been extensively investigated using transient and stable transfection assays with reporter genes [34 - 36]. Distinct elements, including a 16-mer stem-loop that stimulates translation and a 26-mer stem-loop that causes RNA instability in the BF, have been characterized. In addition, GPEET procyclin mRNA possesses a negative regulatory element known as the glycerol responsive element (GRE), which also forms a stem-loop structure and destabilizes GPEET mRNA in late PF [50, 51]. It appears that, by employing several layers of regulation, these parasites can ensure that the rapid changes associated with transmission between the insect vector and the mammalian host are followed by an instant reprogramming of genetic expression.

Studies aimed at understanding the mechanisms controlling gene expression in other trypanosomatids have shown that a significant number of genes are regulated at the mRNA stability level. In contrast to *T. brucei* VSG genes, which are transcribed by RNA pol I, in *Leishmania* and *T. cruzi* there are no reports describing transcriptional regulation affecting the expression of protein-coding genes. Recently, the first direct evidence for epigenetic regulation of RNA pol II transcription in *T. cruzi* was provided by R. Sabatini's group, who described changes in chromatin structure and RNA pol II recruitment to divergent SSRs [13]. By deleting the genes involved in base J synthesis, these authors showed a

global increase in RNA pol II transcription, which was associated with increased histone acetylation and decreased nucleosome abundance. Most of the work describing gene regulation in *T. cruzi* and *Leishmania* has focused on a limited number of gene families that are developmentally regulated during the parasite life cycle including tubulin genes [52], the surface metalloprotease gp63 or *msp* genes [53], the *trans*-sialidase gene family [54] and the amastigote-specific genes encoding A2 [55] and amastin surface proteins [56].

Unlike *T. brucei*, *T. cruzi* and *Leishmania* spp. require an intracellular stage in the mammalian host. *Leishmania* spp. proliferates as promastigotes in the midgut of phlebotomine sand flies and is transmitted to several species of mammals as metacyclic promastigotes when the fly takes a blood meal. In the mammalian host, promastigotes are phagocytized by macrophages, and, once in the phagolysosome, they are converted into amastigotes, which multiply numerous times before being released during cell lysis. *T. cruzi* replicates as epimastigotes in the midgut of different species of reduviid bugs and develops into infective metacyclic trypomastigotes once it reaches the rectum, it is then excreted with the insect feces. Unlike *Leishmania* spp., which multiplies only inside host macrophages, *T. cruzi* trypomastigotes invade essentially any nucleated cell type by a mechanism involving lysosomal recruitment at the parasite invasion site. Also in contrast to *Leishmania*, *T. cruzi* trypomastigotes are able to escape from the phagolysosome into the cytosol, where they differentiate into amastigotes. After several rounds of cell division, amastigotes differentiate again into trypomastigotes and are released from the infected cell and can reach the bloodstream or invade neighboring cells [57]. Details about *Leishmania* and *T. cruzi* life cycles are described in Chapter 1.

Regulatory Elements in the 3' UTR of *T. cruzi* and *Leishmania* Genes

Amastin genes have been a model for several studies aiming at identifying cis and *trans*-acting factors involved in post-transcriptional regulation in *T. cruzi* and *Leishmania*. Initially characterized as *T. cruzi* and *Leishmania* genes whose expression is upregulated in amastigotes, amastin genes have been identified in the genomes of other trypanosomatids, including the insect parasites *Leptomonas seymouri*, *Angomonas deanei*, and *Strigomonas culicis* [57, 58]. Although their

function remains unknown, because of the predicted structure and surface localization in the intracellular stage of *T. cruzi* and *Leishmania* spp., it has been proposed that amastins may play a role in host-parasite interactions within the mammalian cell involved in the transport of ions and nutrients across the membrane or in cell signaling events that trigger parasite differentiation. Phylogenetic analyses of amastin orthologs from various members of the Trypanosomatidae family resulted in a classification of all amastin sequences into four subfamilies, named α-, β-, γ-, and δ-amastins [59]. Importantly, through analyses of *Leishmania* spp., the δ-amastin sub-family has been widely expanded. With 45 copies identified in the genomes of *L. major* and *L. infantum* [60], amastins constitute the largest gene family in the *Leishmania* genus, whose members exhibit regulated expression during the life cycle of the parasite. *T. cruzi*, on the other hand, has only members of the β- and δ-amastins. Similarly to all *Leishmania* amastin genes, *T. cruzi* δ-amastin genes are upregulated in amastigotes compared to those in the epimastigote or trypomastigote stages; however, in contrast to *Leishmania* orthologs, *T. cruzi* β-amastin genes are up-regulated in epimastigotes [61].

The presence of elements in the 3' UTRs plus intergenic regions controlling the expression of amastin genes were examined using transient and stable transfections assays of *T. cruzi* and *Leishmania* with luciferase as a reporter gene [58, 62]. The presence of the δ-amastin 3' UTR plus downstream processing signals of luciferase results in a 36-fold increase in luciferase mRNA levels in *T. cruzi* amastigotes compared to epimastigotes, a difference that corresponds to the 50-fold increase in steady-state levels of amastin mRNA in amastigotes. It was previously shown that this large difference in steady-state amastin mRNA levels between amastigotes and epimastigotes is caused by, at least in part, to differences in amastin mRNA stability because the amastin mRNA is 7-fold more stable in amastigotes than in epimastigotes. Analysis of amastin mRNA in *L. infantum* showed that it is 4-fold more abundant in axenic amastigotes than in promastigotes. The decay kinetics of mRNAs in both stages of the parasite lifecycle indicates the amastin mRNA decays with an overall half-life of aprox. ~40 min in promastigotes *versus* >120 min in amastigotes. Reporter gene assays revealed the presence of a 100 nt long U-rich sequence that contributes

significantly to the stage-specific accumulation and degradation of the amastin mRNA [63]. The 1.8 kb 3′UTR of *L. infantum* amastin mRNA also contains additional cis-acting regulatory sequences related to the *Leishmania*-specific SIDER retroelements [64] and a 300 nt sequence which is highly conserved among a large number of *Leishmania* mRNAs, that confers stage-specific gene expression by a mechanism affecting mRNA translation [63].

Several groups have investigated the presence of regulatory elements in the 3' UTR of other *Leishmania* genes besides amastin genes. The list includes the *Leishmania* aquaglyceroporin AQP1 mRNA [65], HSP70 [66], alfa-tubulin [67], HSP83 [68], histone [69], A2 [55], gp46 and gp63 [70]. A large number of *Leishmania* transcripts harbors in their 3' UTRs two phylogenetically distinct subfamilies of extinct short interspersed degenerated retroposons (SIDER1 and SIDER2) that are involved in post-transcriptional regulation of gene expression [64]. It has been recently shown that members of the SIDER2 subfamily promote mRNA destabilization and that, in contrast to most eukaryotic genes, in which mRNA degradation initiated by poly(A) tail shortening, degradation of SIDER2-containing mRNAs is triggered by site-specific endonucleolytic cleavage within SIDER2 signature sequence, without prior shortening of the poly(A) tail [71].

By conducting stable transfection experiments, we and others have described the role of sequences in the 3’ UTR from various stage-specific and constitutive *T. cruzi* genes that include, in addition to amastin genes, genes encoding α-tubulin and the surface proteins GP72 and GP85 [72, 73]. Using an *in silico* approach, Li and collaborators [74] characterized a 43 nt U-rich RNA element located in the 3' UTR of a large number of *T. cruzi* mRNAs, including *trans*-sialidase, mucin-associated protein (MASP), mucin, surface protease GP63, and protein kinases, that are upregulated in amastigotes. Immunoprecipitation analyses associated with *in silico* genome-wide screening, identified conserved structural RNA motifs, statistically enriched in the 3’ UTR of a group of genes encoding metabolically related transcripts, which are targets for two *T. cruzi* U-rich RNA-binding proteins named TcUBP1 and TcUBP2 [75]. This work not only uncovered signature motifs and conserved structural elements in the 3’ UTR of groups of functionally related mRNAs but also showed that groups of transcripts can be preferentially associated with a given RBP.

RNA-BINDING PROTEINS AS MAJOR PLAYERS OF STAGE-SPECIFIC GENE EXPRESSION IN TRYPANOSOMATIDS

Most studies aimed to understand the mechanisms determining mRNA half-lives and translation efficiency points towards the essential role of RNA-binding proteins (RBPs) and their dynamic interactions with the various subpopulations of mRNA in the different stages of the life cycles of trypanosomatid parasites. Different approaches have been used to identify these RBPs and their mRNA targets. In early studies, after characterizing a cis-acting regulatory element in the 3' UTR, scientists searched for a specific RBP able to recognize this element and influence the expression of the corresponding mRNA using RNA band-shift assays and affinity chromatography protocols. Using this initial gene-based approach, a limited number of RBPs have been characterized in *T. brucei*, *T. cruzi* and *L. infantum*. With the availability of complete genome sequences and, more recently, with new sequencing technologies, not only have much more *cis*-acting RNA regulatory elements been uncovered but many more sequences encoding RBPs have also been identified. Altogether, these studies lead to the description of RNA regulons or RBP-mediated regulatory maps that will significantly improve our understanding of how trypanosome post-transcription regulatory machinery operates. In conjunction with a variety of techniques that include nucleotide UV-cross-linking and RBP-immunoprecipitation coupled to deep sequencing (RIP-seq), high-throughput analyses allow for the characterization of not only many more RBPs containing different RNA binding domains (RBDs) but also detailed analyses of the effects caused by the binding of these regulatory *trans*-acting factors to their target transcripts [76].

Recognition by RBPs of their RNA targets is typically mediated by one or more RBDs, such as the RNA recognition motif (RRM), Cys3His (CCCH) zinc finger, Cold-shock, PIWI and Pumilio-like (PUF) domains (for a recent review, see [77]). Soon after the TriTryp genomes were completed, comparative analyses of predicted RBPs containing the RRM motif identified a total of 139 RRM-containing proteins in *T. cruzi*, 75 in *T. brucei*, and 80 in *L. major* [78]. It should be noted that, because of its hybrid nature, the *T. cruzi* CL Brener genome was assembled and annotated as two redundant datasets of sequences, referred to as "Esmeraldo-like" or "non-Esmeraldo-like" sequences, each corresponding to one

haplotype, which were identified based on the sequences of the CL Brener parental strains [8]. Therefore, a larger number of sequences were identified in the CL Brener genome perhaps because, for various genes, two polymorphic alleles were counted. Interestingly, among the RRM-containing proteins described in all kinetoplastids, only 12 are conserved from kinetoplastids to yeast or mammals, indicating the presence of a large number of trypanosome-specific RRM proteins that may have coevolved with their RNA targets [78]. Furthermore, of the 75 RRM proteins identified in the *T. brucei* genome, only 18 have their function assigned, which include proteins involved in RNA splicing, ribosome biogenesis, translation, and control of mRNA turnover.

Because of the development of an inducible RNAi system, which has enormously facilitated gene function analyses in this parasite, a variety of *T. brucei* RBPs have been functionally characterized. In contrast to *T. cruzi* and *L. major*, which do not have functional RNAi machinery, in *T. brucei*, the efficacy of RNAi knockdown as a tool for systematic analysis of gene function has allowed not only gene-by-gene studies but also high-throughput approaches to studying RBP function (for a recent review, see [79]. Using RNAi screening, C. Clayton's group analyzed the effect of targeting 37 of the 72 RRM-containing proteins of *T. brucei*, showing that knocking down eight of these genes caused clear growth inhibition in BF trypanosomes. Further analysis of one of them, TbRBP3, which was found to be associated with 10 *T. brucei* mRNAs, showed that its depletion did not affect the parasite transcriptome, suggesting a role in translation [80].

Altering the expression levels of RBPs can produce valuable information regarding their function. By overexpressing a single RNA-binding protein, TbRBP6, in cultured noninfectious PF resulted in the appearance of developmental stages observed in the tsetse fly, including metacyclic trypomastigotes expressing a VSG coat [81]. Overexpression or RNAi knockdown of the *T. brucei* RBP named DRBD13 resulted in changes in the expression level of mRNA encoding various membranes associated proteins [82]. Additionally, RIP-seq data suggested that TbRBP6 transcript binds to DRBD13 protein, indicating the involvement of DRBD13 in the differentiation process that occurs in the insect vector and leads to the transformation of PF into the infectious metacyclic trypomastigotes.The first example of an RBP involved in life-cycle stage-specific regulation of the global

trypanosome transcriptome is RBP10, a small RRM-containing protein whose expression is upregulated in BF [83]. Whereas RNAi knockdown of RBP10 in BF resulted in the downregulation of a large number of mRNAs normally found elevated in the bloodstream stage, overexpression of the protein in PF led to an increase in BF-specific mRNAs. Another group of RBPs that have been investigated as factors involved in *T. brucei* development is the ALBA family of proteins (acetylation lowers binding affinity), which are known to be involved in the developmental programs of various microorganisms. *T. brucei* genome encodes four ALBA proteins [84]. Depletion of ALBA3/4 proteins by RNAi in the cultured procyclic stage produces cell modifications mimicking several morphogenetic aspects of trypanosome differentiation that usually occur in the fly midgut [85]. *T. brucei* ALBA proteins have also been characterized as RNA-binding proteins that interact with translation machinery. Initially identified as proteins that interact with elements in the 3' UTR of procyclins, further studies showed that RNAi knockdown of all four Alba proteins specifically reduced translation of a reporter transcript flanked by the GPEET procyclin 3' UTR. Moreover, pull-down assays using tagged Alba proteins revealed interactions with poly(A) binding proteins, ribosomal protein P0 and, in the case of Alba3, the cap-binding protein eIF4E4, further confirming their role in translation control [84].

Recently, ALBA proteins have also been described as *trans*-acting factors regulating amastin developmental gene expression in *Leishmania*. Using RNA affinity chromatography with a 300 nt regulatory region within the amastin 3' UTR as bait, B. Papadopoulou's group identified the homolog of *T. brucei* ALBA3, named LiAlba20, as a specific amastin mRNA-binding factor [86]. Supporting the role of LiAlba20 in amastin gene regulation, gene knockout of LiAlba20 resulted in amastin mRNA destabilization in amastigotes. Further studies have shown that an arginine methyltransferase named LmjPRMT7 is associated with several RBPs, including LmjAlba20. LmjAlba20 is mono-methylated only in LmjPRMT7-expressing promastigote stages, and its protein levels are significantly altered in LmjPRMT7 knockout mutants [87]. In addition to ALBA proteins, only a few other RBPs have been characterized in *Leishmania*. Although proteins containing a PIWI domain such as Argonaute, involved in RNAi machinery, are not found in Old World *Leishmania*, a Piwi-like protein

homolog has been identified in *L. infantum* and *L. major*. Knockout of the gene encoding this Piwi-like protein leads to changes in the expression of several developmental regulated *Leishmania* genes, suggesting that these proteins may play a role in post-transcriptional regulation that is independent of the RNAi pathway [88].

Among the RBPs that have been characterized in *T. cruzi* are several proteins that have been found to be involved in mRNA stabilization and translation. Using polyT beads as bait, S. Goldenberg's group isolated protein complexes bound to poly(A)+ mRNAs from polysomal and post-polysomal fractions of epimastigotes that were in the exponential phase of growth and from epimastigotes under nutritional stress. Mass spectrometry (LC-MS/MS) analyses resulted in the identification of 542 protein components of the mRNP complexes associated with mRNAs [89]. In addition to a number of ribosomal proteins, several RBPs were identified by the authors, including the CCR4-CAF1-NOT proteins implicated in mRNA turnover; the TIA-1-like RBP, known to repress the translation of the mRNAs in response to stress; and Dhh1, which belongs to a highly conserved subfamily of DEAD box helicases known to be involved in global repression of mRNA translation in yeast. These authors also identified TcUBP1, a U-rich RBP containing a single RRM motif, that was previously characterized as an mRNA destabilizing factor that binds to AU-rich elements present in the 3' UTR of mucin mRNAs [90]. Although other studies based on ribosome profiling have shown that control of mRNA translation is more extensive than previously predicted solely by the analyses of total transcript levels [91], the mechanisms involved in this regulation are still far from being understood.

TcUBP1 is the first RBP identified in trypanosomatids as a *trans*-acting factor involved in post-transcriptional control. Together with TcUBP2 and poly-adenylation binding protein, TcUBP1 is part of a ribonucleoprotein (mRNP) complex that recognizes U-rich motifs and is associated with mRNA granules [90]. The analysis of RNA-protein interactions by mRNP immunoprecipitation with specific antibodies against TcUPB1 and TcUPB3 allowed for the identification of 24 mRNAs encoding known proteins, as well as 10 hypothetical genes. The mRNA associated with TcUBP1 was found to encode mucin, amastin and proteases, such as cruzipain and GP63, as well as members of the TS family. In

contrast, the transcripts that interacted with TcUBP3 were found to encode mainly ribosomal proteins [75]. The authors identified a stem-loop structure of 30-35 bases present in the majority of TcUBP-associated mRNAs, which was more frequently represented in the 3' UTRs of mRNAs [75].

RBPs bearing the CCCH zinc finger motif have also been characterized in a *T. cruzi* strain named Dm28c, which has been extensively studied as a model for parasite differentiation [92]. Goldenberg's group showed that the expression of TcZFP1 and TcZFP2 is increased in polysomal fractions from parasites submitted to nutritional stress. TcZFP2 is expressed in amastigotes and trypomastigotes, but its protein levels are reduced in *T. cruzi* metacyclic trypomastigotes. After identifying the TcZFN2 RNA targets using RIP-CHIP, these authors showed that the expression of most genes encoding the TcZFP2-associated mRNAs are upregulated in the metacyclic trypomastigotes, thus suggesting a role for TcZFP2 in metacyclogenesis [93]. TcPUF6 and a few other members of the PUF RNA-binding protein family have also been investigated as potential regulators of metacyclogenesis [94]. Other *T. cruzi* RBPs that have been characterized include two proteins containing an RRM motif: TcRBP19, which was only detected in the amastigote stage [95], and TcRBP40, which is a *T. cruzi*-specific RBP whose mRNA targets mainly encode putative *T. cruzi* transmembrane proteins [96].

Clearly, the number of RBPs known to be involved in post-transcriptional regulation in trypanosomatids is growing at a faster rate than it was a few years ago. However, the exact mechanism employed by these parasites to control the expression of their genes, including the ones coding for these RBPs, are far from being understood. Although almost all the main trypanosome homologs of yeast/mammalian mRNA degradation enzymes (such as components of the exosome), decapping and deadenylating enzymes, as well as exoribonucleases have been characterized and major pathways identified, there is still much work that must be done to understand how these activities are regulated. Because of the emerging role of non-coding RNA, which is now clearly established in various eukaryotes, several groups have also investigated whether small non-coding RNAs also participate in the regulatory processes associated with parasite differentiation [97]. However, although various components of the small RNome have been characterized in all three groups of parasites, to date, no clear evidence

exists for the involvement of these molecules in gene regulation processes. With the powerful genome-wide approaches that have recently become available, the main gene regulatory networks, together with the RNA and protein components involved in these networks, will soon be fully characterized. Because they are essential components of the survival arsenal employed by these parasites, a better understating of these regulatory machines will ultimately reveal more adequate targets for pharmacological interventions that can block parasite development and lead to disease control.

CONFLICT OF INTEREST

The authors confirm that they have no conflict of interest to declare for this publication.

ACKNOWLEDGEMENTS

The work in our laboratory was supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) and from the Instituto Nacional de Ciência e Tecnologia de Vacinas (INCTV).

LIST OF ABBREVIATIONS

BF	Bloodstream form
ES	Expression site
ESB	Expression site body
kDNA	Kinetoplast DNA
PF	Procyclic form
PPT	Polypyrimidine tract
PTU	Polycistronic transcription unit
PUF	Pumilio-like domain
RBD	RNA binding domain
RBP	RNA binding protein
RNAi	RNA interference
RNA pol I	RNA polymerase I

RNA Pol II RNA polymerase II

RRM RNA recognition motif

SL Spliced leader

UTR Untranslated region

VSG Variant surface glycoprotein

REFERENCES

[1] Fernandes AP, Nelson K, Beverley SM. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism. Proc Natl Acad Sci USA 1993; 90(24): 11608-12.
[http://dx.doi.org/10.1073/pnas.90.24.11608] [PMID: 8265597]

[2] Michaeli S. Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. Future Microbiol 2011; 6(4): 459-74.
[http://dx.doi.org/10.2217/fmb.11.20] [PMID: 21526946]

[3] Teixeira SM, daRocha WD. Control of gene expression and genetic manipulation in the Trypanosomatidae. Genet Mol Res 2003; 2(1): 148-58.
[PMID: 12917811]

[4] Lukes J, Hashimi H, Zíková A. Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. Curr Genet 2005; 48(5): 277-99.
[http://dx.doi.org/10.1007/s00294-005-0027-0] [PMID: 16215758]

[5] Myler PJ, Audleman L, deVos T, *et al. Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. Proc Natl Acad Sci USA 1999; 96(6): 2902-6.
[http://dx.doi.org/10.1073/pnas.96.6.2902] [PMID: 10077609]

[6] Thomas S, Green A, Sturm NR, Campbell DA, Myler PJ. Histone acetylations mark origins of polycistronic transcription in *Leishmania major.* BMC Genomics 2009; 10: 152.
[http://dx.doi.org/10.1186/1471-2164-10-152] [PMID: 19356248]

[7] Berriman M, Ghedin E, Hertz-Fowler C, *et al.* The genome of the African trypanosome *Trypanosoma brucei.* Science 2005; 309(5733): 416-22.
[http://dx.doi.org/10.1126/science.1112642] [PMID: 16020726]

[8] El-Sayed NM, Myler PJ, Bartholomeu DC, *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 2005; 309(5733): 409-15.
[http://dx.doi.org/10.1126/science.1112631] [PMID: 16020725]

[9] Ivens AC, Peacock CS, Worthey EA, *et al.* The genome of the kinetoplastid parasite, *Leishmania major.* Science 2005; 309(5733): 436-42.
[http://dx.doi.org/10.1126/science.1112680] [PMID: 16020728]

[10] Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. PLoS Pathog 2010; 6(9): e1001090.
[http://dx.doi.org/10.1371/journal.ppat.1001090] [PMID: 20838601]

[11] Martinez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martinez LE, *et al.* Gene expression in trypanosomatid parasites. J Biomed Biotechnol 2010; 2010: 525241.

[12] Cliffe LJ, Siegel TN, Marshall M, Cross GA, Sabatini R. Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of *Trypanosoma brucei.* Nucleic Acids Res 2010; 38(12): 3923-35.
[http://dx.doi.org/10.1093/nar/gkq146] [PMID: 20215442]

[13] Ekanayake D, Sabatini R. Epigenetic regulation of polymerase II transcription initiation in Trypanosoma cruzi: modulation of nucleosome abundance, histone modification, and polymerase occupancy by O-linked thymine DNA glucosylation. Eukaryot Cell 2011; 10(11): 1465-72.
[http://dx.doi.org/10.1128/EC.05185-11] [PMID: 21926332]

[14] van Luenen HG, Farris C, Jan S, *et al.* Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in *Leishmania.* Cell 2012; 150(5): 909-21.
[http://dx.doi.org/10.1016/j.cell.2012.07.030] [PMID: 22939620]

[15] Donelson JE, Zeng W. A comparison of *trans*-RNA splicing in trypanosomes and nematodes. Parasitol Today (Regul Ed) 1990; 6(10): 327-34.
[http://dx.doi.org/10.1016/0169-4758(90)90177-6] [PMID: 15463258]

[16] LeBowitz JH, Smith HQ, Rusche L, Beverley SM. Coupling of poly(A) site selection and *trans*-splicing in *Leishmania.* Genes Dev 1993; 7(6): 996-1007.
[http://dx.doi.org/10.1101/gad.7.6.996] [PMID: 8504937]

[17] Matthews KR, Tschudi C, Ullu E. A common pyrimidine-rich motif governs *trans*-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes Dev 1994; 8(4): 491-501.
[http://dx.doi.org/10.1101/gad.8.4.491] [PMID: 7907303]

[18] Schürch N, Hehl A, Vassella E, Braun R, Roditi I. Accurate polyadenylation of procyclin mRNAs in *Trypanosoma brucei* is determined by pyrimidine-rich elements in the intergenic regions. Mol Cell Biol 1994; 14(6): 3668-75.
[http://dx.doi.org/10.1128/MCB.14.6.3668] [PMID: 7910942]

[19] Liang XH, Haritan A, Uliel S, Michaeli S. trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot Cell 2003; 2(5): 830-40.
[http://dx.doi.org/10.1128/EC.2.5.830-840.2003] [PMID: 14555465]

[20] Siegel TN, Hekstra DR, Wang X, Dewell S, Cross GA. Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. Nucleic Acids Res 2010; 38(15): 4946-57.
[http://dx.doi.org/10.1093/nar/gkq237] [PMID: 20385579]

[21] Nilsson D, Gunasekera K, Mani J, *et al.* Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei.* PLoS Pathog 2010; 6(8): e1001037.
[http://dx.doi.org/10.1371/journal.ppat.1001037] [PMID: 20700444]

[22] Dillon LA, Okrah K, Hughitt VK, *et al.* Transcriptomic profiling of gene expression and RNA processing during *Leishmania major* differentiation. Nucleic Acids Res 2015; 43(14): 6799-813.

[http://dx.doi.org/10.1093/nar/gkv656] [PMID: 26150419]

[23] Campos PC, Bartholomeu DC, DaRocha WD, Cerqueira GC, Teixeira SM. Sequences involved in mRNA processing in *Trypanosoma cruzi.* Int J Parasitol 2008; 38(12): 1383-9.
[http://dx.doi.org/10.1016/j.ijpara.2008.07.001] [PMID: 18700146]

[24] El-Sayed NM, Myler PJ, Blandin G, *et al.* Comparative genomics of trypanosomatid parasitic protozoa. Science 2005; 309(5733): 404-9.
[http://dx.doi.org/10.1126/science.1112181] [PMID: 16020724]

[25] Clayton CE. Networks of gene expression regulation in *Trypanosoma brucei.* Mol Biochem Parasitol 2014; 195(2): 96-106.
[http://dx.doi.org/10.1016/j.molbiopara.2014.06.005] [PMID: 24995711]

[26] De Gaudenzi JG, Carmona SJ, Agüero F, Frasch AC. Genome-wide analysis of 3-untranslated regions supports the existence of post-transcriptional regulons controlling gene expression in trypanosomes. Peer J 2013; 1: e118.
[http://dx.doi.org/10.7717/peerj.118] [PMID: 23904995]

[27] Moretti NS, Schenkman S. Chromatin modifications in trypanosomes due to stress. Cell Microbiol 2013; 15(5): 709-17.
[http://dx.doi.org/10.1111/cmi.12111] [PMID: 23336291]

[28] Horn D, McCulloch R. Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. Curr Opin Microbiol 2010; 13(6): 700-5.
[http://dx.doi.org/10.1016/j.mib.2010.08.009] [PMID: 20884281]

[29] Vassella E, Acosta-Serrano A, Studer E, Lee SH, Englund PT, Roditi I. Multiple procyclin isoforms are expressed differentially during the development of insect forms of *Trypanosoma brucei.* J Mol Biol 2001; 312(4): 597-607.
[http://dx.doi.org/10.1006/jmbi.2001.5004] [PMID: 11575917]

[30] Ross R, Thomson D. A case of sleeping sickness showing regular periodical increase of the parasites disclosed. BMJ 1910; 1(2582): 1544-5.
[http://dx.doi.org/10.1136/bmj.1.2582.1544] [PMID: 20765166]

[31] Cross GA. Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei.* Parasitology 1975; 71(3): 393-417.
[http://dx.doi.org/10.1017/S003118200004717X] [PMID: 645]

[32] Bridgen PJ, Cross GA, Bridgen J. N-terminal amino acid sequences of variant-specific surface antigens from *Trypanosoma brucei.* Nature 1976; 263(5578): 613-4.
[http://dx.doi.org/10.1038/263613a0] [PMID: 980109]

[33] Williams RO, Young JR, Majiwa PA. Genomic rearrangements correlated with antigenic variation in *Trypanosoma brucei.* Nature 1979; 282(5741): 847-9.
[http://dx.doi.org/10.1038/282847a0] [PMID: 514362]

[34] Kooter JM, van der Spek HJ, Wagter R, *et al.* The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in *T. brucei.* Cell 1987; 51(2): 261-72.
[http://dx.doi.org/10.1016/0092-8674(87)90153-X] [PMID: 2444341]

[35] Navarro M, Gull K. A pol I transcriptional body associated with VSG mono-allelic expression in

Trypanosoma brucei. Nature 2001; 414(6865): 759-63.
[http://dx.doi.org/10.1038/414759a] [PMID: 11742402]

[36] Povelones ML, Gluenz E, Dembek M, Gull K, Rudenko G. Histone H1 plays a role in heterochromatin formation and VSG expression site silencing in *Trypanosoma brucei.* PLoS Pathog 2012; 8(11): e1003010.
[http://dx.doi.org/10.1371/journal.ppat.1003010] [PMID: 23133390]

[37] Wang QP, Kawahara T, Horn D. Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes. Mol Microbiol 2010; 77(5): 1237-45.
[http://dx.doi.org/10.1111/j.1365-2958.2010.07284.x] [PMID: 20624217]

[38] Günzl A, Kirkham JK, Nguyen TN, Badjatia N, Park SH. Mono-allelic VSG expression by RNA polymerase I in *Trypanosoma brucei*: expression site control from both ends? Gene 2015; 556(1): 68-73.
[http://dx.doi.org/10.1016/j.gene.2014.09.047] [PMID: 25261847]

[39] Glover L, Horn D. Locus-specific control of DNA resection and suppression of subtelomeric VSG recombination by HAT3 in the African trypanosome. Nucleic Acids Res 2014; 42(20): 12600-13.
[http://dx.doi.org/10.1093/nar/gku900] [PMID: 25300492]

[40] Glover L, Hutchinson S, Alsford S, McCulloch R, Field MC, Horn D. Antigenic variation in African trypanosomes: the importance of chromosomal and nuclear context in VSG expression control. Cell Microbiol 2013; 15(12): 1984-93.
[http://dx.doi.org/10.1111/cmi.12215] [PMID: 24047558]

[41] van Leeuwen F, Taylor MC, Mondragon A, *et al.* beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. Proc Natl Acad Sci USA 1998; 95(5): 2366-71.
[http://dx.doi.org/10.1073/pnas.95.5.2366] [PMID: 9482891]

[42] Mony BM, MacGregor P, Ivens A, *et al.* Genome-wide dissection of the quorum sensing signalling pathway in *Trypanosoma brucei.* Nature 2014; 505(7485): 681-5.
[http://dx.doi.org/10.1038/nature12864] [PMID: 24336212]

[43] Cirovic O, Ochsenreiter T. Whole proteome analysis of the protozoan parasite *Trypanosoma brucei* using stable isotope labeling by amino acids in cell culture and mass spectrometry. Methods Mol Biol 2014; 1188: 47-55.
[http://dx.doi.org/10.1007/978-1-4939-1142-4_4] [PMID: 25059603]

[44] Roditi I. The VSG-procyclin switch. Parasitol Today (Regul Ed) 1996; 12(2): 47-9.
[http://dx.doi.org/10.1016/0169-4758(96)80651-X] [PMID: 15275251]

[45] Roditi I, Furger A, Ruepp S, Schürch N, Bütikofer P. Unravelling the procyclin coat of *Trypanosoma brucei.* Mol Biochem Parasitol 1998; 91(1): 117-30.
[http://dx.doi.org/10.1016/S0166-6851(97)00195-3] [PMID: 9574930]

[46] Haenni S, Studer E, Burkard GS, Roditi I. Bidirectional silencing of RNA polymerase I transcription by a strand switch region in *Trypanosoma brucei.* Nucleic Acids Res 2009; 37(15): 5007-18.
[http://dx.doi.org/10.1093/nar/gkp513] [PMID: 19531741]

[47] Rudenko G, Le Blancq S, Smith J, Lee MG, Rattray A, Van der Ploeg LH. Procyclic acidic repetitive

protein (PARP) genes located in an unusually small alpha-amanitin-resistant transcription unit: PARP promoter activity assayed by transient DNA transfection of *Trypanosoma brucei.* Mol Cell Biol 1990; 10(7): 3492-504.
[http://dx.doi.org/10.1128/MCB.10.7.3492] [PMID: 1694012]

[48] Pays E, Coquelet H, Tebabi P, *et al. Trypanosoma brucei*: constitutive activity of the VSG and procyclin gene promoters. EMBO J 1990; 9(10): 3145-51.
[PMID: 1698609]

[49] Knüsel S, Roditi I. Insights into the regulation of GPEET procyclin during differentiation from early to late procyclic forms of *Trypanosoma brucei.* Mol Biochem Parasitol 2013; 191(2): 66-74.
[http://dx.doi.org/10.1016/j.molbiopara.2013.09.004] [PMID: 24076427]

[50] Furger A, Schürch N, Kurath U, Roditi I. Elements in the 3 untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation. Mol Cell Biol 1997; 17(8): 4372-80.
[http://dx.doi.org/10.1128/MCB.17.8.4372] [PMID: 9234695]

[51] Hehl A, Vassella E, Braun R, Roditi I. A conserved stem-loop structure in the 3 untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei.* Proc Natl Acad Sci USA 1994; 91(1): 370-4.
[http://dx.doi.org/10.1073/pnas.91.1.370] [PMID: 8278396]

[52] Coulson RM, Connor V, Chen JC, Ajioka JW. Differential expression of *Leishmania major* beta-tubulin genes during the acquisition of promastigote infectivity. Mol Biochem Parasitol 1996; 82(2): 227-36.
[http://dx.doi.org/10.1016/0166-6851(96)02739-9] [PMID: 8946388]

[53] Ramamoorthy R, Swihart KG, McCoy JJ, Wilson ME, Donelson JE. Intergenic regions between tandem gp63 genes influence the differential expression of gp63 RNAs in *Leishmania chagasi* promastigotes. J Biol Chem 1995; 270(20): 12133-9.
[http://dx.doi.org/10.1074/jbc.270.20.12133] [PMID: 7744862]

[54] Jäger AV, Muiá RP, Campetella O. Stage-specific expression of *Trypanosoma cruzi trans*-sialidase involves highly conserved 3 untranslated regions. FEMS Microbiol Lett 2008; 283(2): 182-8.
[http://dx.doi.org/10.1111/j.1574-6968.2008.01170.x] [PMID: 18422619]

[55] Charest H, Zhang WW, Matlashewski G. The developmental expression of *Leishmania donovani* A2 amastigote-specific genes is post-transcriptionally mediated and involves elements located in the 3-untranslated region. J Biol Chem 1996; 271(29): 17081-90.
[http://dx.doi.org/10.1074/jbc.271.29.17081] [PMID: 8663340]

[56] Teixeira SM, Russell DG, Kirchhoff LV, Donelson JE. A differentially expressed gene family encoding amastin, a surface protein of *Trypanosoma cruzi* amastigotes. J Biol Chem 1994; 269(32): 20509-16.
[PMID: 8051148]

[57] Bartholomeu DC, de Paiva RM, Mendes TA, DaRocha WD, Teixeira SM. Unveiling the intracellular survival gene kit of trypanosomatid parasites. PLoS Pathog 2014; 10(12): e1004399.
[http://dx.doi.org/10.1371/journal.ppat.1004399] [PMID: 25474314]

[58] Teixeira SM, Kirchhoff LV, Donelson JE. Post-transcriptional elements regulating expression of

mRNAs from the amastin/tuzin gene cluster of *Trypanosoma cruzi*. J Biol Chem 1995; 270(38): 22586-94.
[http://dx.doi.org/10.1074/jbc.270.38.22586] [PMID: 7673251]

[59] Jackson AP. The evolution of amastin surface glycoproteins in trypanosomatid parasites. Mol Biol Evol 2010; 27(1): 33-45.
[http://dx.doi.org/10.1093/molbev/msp214] [PMID: 19748930]

[60] Rochette A, McNicoll F, Girard J, *et al.* Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in *Leishmania* spp. Mol Biochem Parasitol 2005; 140(2): 205-20.
[http://dx.doi.org/10.1016/j.molbiopara.2005.01.006] [PMID: 15760660]

[61] Kangussu-Marcolino MM, de Paiva RM, Araújo PR, *et al.* Distinct genomic organization, mRNA expression and cellular localization of members of two amastin sub-families present in *Trypanosoma cruzi*. BMC Microbiol 2013; 13: 10.
[http://dx.doi.org/10.1186/1471-2180-13-10] [PMID: 23327097]

[62] Coughlin BC, Teixeira SM, Kirchhoff LV, Donelson JE. Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3-untranslated region position-dependent cis-element and an untranslated region-binding protein. J Biol Chem 2000; 275(16): 12051-60.
[http://dx.doi.org/10.1074/jbc.275.16.12051] [PMID: 10766837]

[63] Boucher N, Wu Y, Dumas C, *et al.* A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3-untranslated region element. J Biol Chem 2002; 277(22): 19511-20.
[http://dx.doi.org/10.1074/jbc.M200500200] [PMID: 11912202]

[64] Bringaud F, Müller M, Cerqueira GC, *et al.* Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania.* PLoS Pathog 2007; 3(9): 1291-307.
[http://dx.doi.org/10.1371/journal.ppat.0030136] [PMID: 17907803]

[65] Mandal G, Mandal S, Sharma M, *et al.* Species-specific antimonial sensitivity in Leishmania is driven by post-transcriptional regulation of AQP1. PLoS Negl Trop Dis 2015; 9(2): e0003500.
[http://dx.doi.org/10.1371/journal.pntd.0003500] [PMID: 25714343]

[66] Requena JM, Chicharro C, García L, Parrado R, Puerta CJ, Cañavate C. Sequence analysis of the 3-untranslated region of HSP70 (type I) genes in the genus *Leishmania*: its usefulness as a molecular marker for species identification. Parasit Vectors 2012; 5: 87.
[http://dx.doi.org/10.1186/1756-3305-5-87] [PMID: 22541251]

[67] Ramírez CA, Requena JM, Puerta CJ. Alpha tubulin genes from *Leishmania braziliensis*: genomic organization, gene structure and insights on their expression. BMC Genomics 2013; 14: 454.
[http://dx.doi.org/10.1186/1471-2164-14-454] [PMID: 23829570]

[68] David M, Gabdank I, Ben-David M, *et al.* Preferential translation of Hsp83 in Leishmania requires a thermosensitive polypyrimidine-rich element in the 3 UTR and involves scanning of the 5 UTR. RNA 2010; 16(2): 364-74.
[http://dx.doi.org/10.1261/rna.1874710] [PMID: 20040590]

[69] Abanades DR, Ramírez L, Iborra S, *et al.* Key role of the 3 untranslated region in the cell cycle regulated expression of the *Leishmania infantum* histone H2A genes: minor synergistic effect of the 5

untranslated region. BMC Mol Biol 2009; 10: 48.
[http://dx.doi.org/10.1186/1471-2199-10-48] [PMID: 19460148]

[70] Myung KS, Beetham JK, Wilson ME, Donelson JE. Comparison of the post-transcriptional regulation of the mRNAs for the surface proteins PSA (GP46) and MSP (GP63) of *Leishmania chagasi.* J Biol Chem 2002; 277(19): 16489-97.
[http://dx.doi.org/10.1074/jbc.M200174200] [PMID: 11856749]

[71] Haile S, Dupé A, Papadopoulou B. Deadenylation-independent stage-specific mRNA degradation in *Leishmania.* Nucleic Acids Res 2008; 36(5): 1634-44.
[http://dx.doi.org/10.1093/nar/gkn019] [PMID: 18250085]

[72] Araújo PR, Teixeira SM. Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi*: a review. Mem Inst Oswaldo Cruz 2011; 106(3): 257-66.
[http://dx.doi.org/10.1590/S0074-02762011000300002] [PMID: 21655811]

[73] Nozaki T, Cross GA. Effects of 3 untranslated and intergenic regions on gene expression in *Trypanosoma cruzi.* Mol Biochem Parasitol 1995; 75(1): 55-67.
[http://dx.doi.org/10.1016/0166-6851(95)02512-X] [PMID: 8720175]

[74] Li ZH, De Gaudenzi JG, Alvarez VE, *et al.* A 43-nucleotide U-rich element in 3-untranslated region of large number of *Trypanosoma cruzi* transcripts is important for mRNA abundance in intracellular amastigotes. J Biol Chem 2012; 287(23): 19058-69.
[http://dx.doi.org/10.1074/jbc.M111.338699] [PMID: 22500021]

[75] Noé G, De Gaudenzi JG, Frasch AC. Functionally related transcripts have common RNA motifs for specific RNA-binding proteins in trypanosomes. BMC Mol Biol 2008; 9: 107.
[http://dx.doi.org/10.1186/1471-2199-9-107] [PMID: 19063746]

[76] DOrso I, De Gaudenzi JG, Frasch AC. RNA-binding proteins and mRNA turnover in trypanosomes. Trends Parasitol 2003; 19(4): 151-5.
[http://dx.doi.org/10.1016/S1471-4922(03)00035-7] [PMID: 12689640]

[77] Sanchez-Diaz P, Penalva LO. Post-transcription meets post-genomic: the saga of RNA binding proteins in a new era. RNA Biol 2006; 3(3): 101-9.
[http://dx.doi.org/10.4161/rna.3.3.3373] [PMID: 17114948]

[78] De Gaudenzi J, Frasch AC, Clayton C. RNA-binding domain proteins in Kinetoplastids: a comparative analysis. Eukaryot Cell 2005; 4(12): 2106-14.
[http://dx.doi.org/10.1128/EC.4.12.2106-2114.2005] [PMID: 16339728]

[79] Kolev NG, Ullu E, Tschudi C. The emerging role of RNA-binding proteins in the life cycle of *Trypanosoma brucei.* Cell Microbiol 2014; 16(4): 482-9.
[http://dx.doi.org/10.1111/cmi.12268] [PMID: 24438230]

[80] Wurst M, Robles A, Po J, *et al.* An RNAi screen of the RRM-domain proteins of *Trypanosoma brucei.* Mol Biochem Parasitol 2009; 163(1): 61-5.
[http://dx.doi.org/10.1016/j.molbiopara.2008.09.001] [PMID: 18840477]

[81] Kolev NG, Ramey-Butler K, Cross GA, Ullu E, Tschudi C. Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. Science 2012; 338(6112): 1352-3.

[http://dx.doi.org/10.1126/science.1229641] [PMID: 23224556]

[82] Jha BA, Gazestani VH, Yip CW, Salavati R. The DRBD13 RNA binding protein is involved in the insect-stage differentiation process of *Trypanosoma brucei.* FEBS Lett 2015; 589(15): 1966-74. [http://dx.doi.org/10.1016/j.febslet.2015.05.036] [PMID: 26028502]

[83] Wurst M, Seliger B, Jha BA, Klein C, Queiroz R, Clayton C. Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in Trypanosoma brucei. Mol Microbiol 2012; 83(5): 1048-63. [http://dx.doi.org/10.1111/j.1365-2958.2012.07988.x] [PMID: 22296558]

[84] Mani J, Güttinger A, Schimanski B, *et al.* Alba-domain proteins of *Trypanosoma brucei* are cytoplasmic RNA-binding proteins that interact with the translation machinery. PLoS One 2011; 6(7): e22463. [http://dx.doi.org/10.1371/journal.pone.0022463] [PMID: 21811616]

[85] Subota I, Rotureau B, Blisnick T, *et al.* ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation. Mol Biol Cell 2011; 22(22): 4205-19. [http://dx.doi.org/10.1091/mbc.E11-06-0511] [PMID: 21965287]

[86] Dupé A, Dumas C, Papadopoulou B. An Alba-domain protein contributes to the stage-regulated stability of amastin transcripts in *Leishmania.* Mol Microbiol 2014; 91(3): 548-61. [http://dx.doi.org/10.1111/mmi.12478] [PMID: 24620725]

[87] Ferreira TR, Alves-Ferreira EV, Defina TP, Walrad P, Papadopoulou B, Cruz AK. Altered expression of an RBP-associated arginine methyltransferase 7 in *Leishmania major* affects parasite infection. Mol Microbiol 2014. [http://dx.doi.org/10.1111/mmi.12819] [PMID: 25294169]

[88] Padmanabhan PK, Dumas C, Samant M, Rochette A, Simard MJ, Papadopoulou B. Novel features of a PIWI-like protein homolog in the parasitic protozoan *Leishmania.* PLoS One 2012; 7(12): e52612. [http://dx.doi.org/10.1371/journal.pone.0052612] [PMID: 23285111]

[89] Alves LR, Avila AR, Correa A, *et al.* Proteomic analysis reveals the dynamic association of proteins with translated mRNAs in *Trypanosoma cruzi.* Gene 2010; 452(2): 72-8. [http://dx.doi.org/10.1016/j.gene.2009.12.009] [PMID: 20060445]

[90] DOrso I, Frasch AC. TcUBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element- and Poly(A)-binding proteins forming a ribonucleoprotein complex. J Biol Chem 2002; 277(52): 50520-8. [http://dx.doi.org/10.1074/jbc.M209092200] [PMID: 12403777]

[91] Jensen BC, Ramasamy G, Vasconcelos EJ, Ingolia NT, Myler PJ, Parsons M. Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei.* BMC Genomics 2014; 15: 911. [http://dx.doi.org/10.1186/1471-2164-15-911] [PMID: 25331479]

[92] Alves LR, Oliveira C, Mörking PA, *et al.* The mRNAs associated to a zinc finger protein from *Trypanosoma cruzi* shift during stress conditions. RNA Biol 2014; 11(7): 921-33. [http://dx.doi.org/10.4161/rna.29622] [PMID: 25180711]

[93] Mörking PA, Rampazzo RdeC, Walrad P, *et al.* The zinc finger protein TcZFP2 binds target mRNAs

enriched during *Trypanosoma cruzi* metacyclogenesis. Mem Inst Oswaldo Cruz 2012; 107(6): 790-9. [http://dx.doi.org/10.1590/S0074-02762012000600014] [PMID: 22990970]

[94] Dallagiovanna B, Pérez L, Sotelo-Silveira J, Smircich P, Duhagon MA, Garat B. *Trypanosoma cruzi*: molecular characterization of TcPUF6, a Pumilio protein. Exp Parasitol 2005; 109(4): 260-4. [http://dx.doi.org/10.1016/j.exppara.2005.01.003] [PMID: 15755425]

[95] Pérez-Díaz L, Correa A, Moretão MP, Goldenberg S, Dallagiovanna B, Garat B. The overexpression of the trypanosomatid-exclusive TcRBP19 RNA-binding protein affects cellular infection by *Trypanosoma cruzi*. Mem Inst Oswaldo Cruz 2012; 107(8): 1076-9. [http://dx.doi.org/10.1590/S0074-02762012000800021] [PMID: 23295764]

[96] Guerra-Slompo EP, Probst CM, Pavoni DP, Goldenberg S, Krieger MA, Dallagiovanna B. Molecular characterization of the *Trypanosoma cruzi* specific RNA binding protein TcRBP40 and its associated mRNAs. Biochem Biophys Res Commun 2012; 420(2): 302-7. [http://dx.doi.org/10.1016/j.bbrc.2012.02.154] [PMID: 22425988]

[97] Michaeli S. Non-coding RNA and the complex regulation of the trypanosome life cycle. Curr Opin Microbiol 2014; 20: 146-52. [http://dx.doi.org/10.1016/j.mib.2014.06.006] [PMID: 25063970]

Apêndice III: Paper publicado na revista Plos Pathogens com dados de expressão diferencial entre os clones CL Brener e CL-14 do *T. cruzi*

PLOS | PATHOGENS

RESEARCH ARTICLE

Comparative transcriptome profiling of virulent and non-virulent *Trypanosoma cruzi* underlines the role of surface proteins during infection

A. Trey Belew[1☯], Caroline Junqueira[2☯], Gabriela F. Rodrigues-Luiz[3☯], Bruna M. Valente[4], Antonio Edson R. Oliveira[4], Rafael B. Polidoro[5], Luciana W. Zuccherato[4], Daniella C. Bartholomeu[3], Sergio Schenkman[5], Ricardo T. Gazzinelli[2,4], Barbara A. Burleigh[6], Najib M. El-Sayed[1]*, Santuza M. R. Teixeira[4]*

1 Department of Cell Biology and Molecular Genetics and Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland, United States of America, **2** Centro de Pesquisas Rene Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Minas Gerais, Brazil, **3** Departamento de Parasitologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, **4** Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, **5** Departmento de Microbiologia, Imunologia e Parasitologia, Universidade Federal de São Paulo, São Paulo, São Paulo, Brazil, **6** Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States of America

☯ These authors contributed equally to this work.
* santuzat@ufmg.br (SMRT); elsayed@umd.edu (NES)

Citation: Belew AT, Junqueira C, Rodrigues-Luiz GF, Valente BM, Oliveira AER, Polidoro RB, et al. (2017) Comparative transcriptome profiling of virulent and non-virulent *Trypanosoma cruzi* underlines the role of surface proteins during infection. PLoS Pathog 13(12): e1006767. https://doi.org/10.1371/journal.ppat.1006767

Editor: Martin Taylor, London School of Hygiene & Tropical Medicine, UNITED STATES

Received: June 17, 2017

Accepted: November 22, 2017

Published: December 14, 2017

Data Availability Statement: RNA-Seq data were deposited at the National Center for Biotechnology (NCBI) Sequence Read Archive (SRA) under bioprojects PRJNA389926 (for uninfected and infected HFFs) and bioproject PRJNA389925 (for data from T. cruzi trypomastigotes).

Funding: SMRT is supported by funds from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, 445456/2014-0), Fundação de Amparo a Pesquisa do Estado de Minas Gerais

Abstract

Trypanosoma cruzi, the protozoan that causes Chagas disease, has a complex life cycle involving several morphologically and biochemically distinct stages that establish intricate interactions with various insect and mammalian hosts. It has also a heterogeneous population structure comprising strains with distinct properties such as virulence, sensitivity to drugs, antigenic profile and tissue tropism. We present a comparative transcriptome analysis of two cloned *T. cruzi* strains that display contrasting virulence phenotypes in animal models of infection: CL Brener is a virulent clone and CL-14 is a clone that is neither infective nor pathogenic in *in vivo* models of infection. Gene expression analysis of trypomastigotes and intracellular amastigotes harvested at 60 and 96 hours post-infection (hpi) of human fibroblasts revealed large differences that reflect the parasite's adaptation to distinct environments during the infection of mammalian cells, including changes in energy sources, oxidative stress responses, cell cycle control and cell surface components. While extensive transcriptome remodeling was observed when trypomastigotes of both strains were compared to 60 hpi amastigotes, differences in gene expression were much less pronounced when 96 hpi amastigotes and trypomastigotes of CL Brener were compared. In contrast, the differentiation of the avirulent CL-14 from 96 hpi amastigotes to extracellular trypomastigotes was associated with considerable changes in gene expression, particularly in gene families encoding surface proteins such as trans-sialidases, mucins and the mucin associated surface proteins (MASPs). Thus, our comparative transcriptome analysis indicates that the avirulent phenotype of CL-14 may be due, at least in part, to a reduced or delayed

(FAPEMIG, APQ-00805-15), Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior (CAPES) and the Instituto Nacional de Ciencia e Tecnologia de Vacinas (INCTV). ATB and NES were supported in part by a grant from the NIH to NES (AI094773). BAB and NES were supported in part by a grant from the NIH (AI094195) to BAB and NES. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

expression of genes encoding surface proteins that are associated with the transition of amastigotes to trypomastigotes, an essential step in the establishment of the infection in the mammalian host. Confirming the role of members of the trans-sialidase family of surface proteins for parasite differentiation, transfected CL-14 constitutively expressing a trans-sialidase gene displayed faster kinetics of trypomastigote release in the supernatant of infected cells compared to wild type CL-14.

Author summary

Chagas disease, caused by the protozoan parasite *Trypanosoma cruzi*, is an infection that occurs in several Latin American countries, resulting in a mild illness or in severe damage of the heart and intestinal tract. Such a broad spectrum of clinical manifestations observed in Chagas disease patients is likely due to differences in host susceptibility as well as to a large heterogeneity among *T*. *cruzi* isolates. The identification of virulence factors that are differentially expressed in the parasite population is a valuable strategy for understanding of the distinct interactions that occur between this pathogen and its host, which may or may not lead to pathogenesis. By comparing the gene expression profiles of two *T*. *cruzi* strains that display contrasting virulence phenotypes in animal models of infection, we identified a central role for genes encoding surface proteins that is associated with the differentiation from intracellular replicative amastigotes to infective trypomastigotes. We showed that the expression of these genes occurs differentially within the two strains and this difference may be a factor that impacts parasite survival and dissemination in the mammalian host.

Introduction

Trypanosoma cruzi is the etiological agent of Chagas disease, a disease affecting at least 8 million people throughout Latin America and for which there are only two drugs available, both with poor efficacy and harmful side effects (http://www.who.int/mediacentre/factsheets/fs340/en/). *T*. *cruzi* infection begins with metacyclic trypomastigotes released in the feces of a triatominae vector during a blood meal. After reaching the bloodstream through skin cuts or through the mucosa, trypomastigotes invade different cell types in the mammalian host. Once in the cytoplasm, they differentiate into replicative non-flagellated amastigotes, which undergo several rounds of division before differentiating again into trypomastigotes and bursting out of the host cell. Bloodstream trypomastigotes, ingested by the vector during a blood meal, differentiate into epimastigotes and replicate in the insect midgut [1]. Such a complex life cycle, involving several morphologically and biochemically distinct parasite forms and very different host environments, requires complex regulatory mechanisms to direct gene expression and the adaptive strategies that allow parasites to thrive within distinct hosts.

The sequence of the *T*. *cruzi* CL Brener genome, with an estimated haploid size of 55 Mb, revealed a highly repetitive genome with protein coding genes organized in long, unidirectional polycistronic transcription units [2]. Because of its hybrid nature and repetitive content, which prevented its complete assembly, the CL Brener genome is represented by two datasets of contigs, each corresponding to one haplotype [2,3]. The two haplotypes are referred to as "esmeraldo-like" or "non-esmeraldo-like" based on their parental strains [2,4]. Several genes,

particularly members of large multicopy gene families, were excluded from the current assembly of the CL Brener genome, which is predicted to contain 41 pairs of chromosomes [3].

The naturally occurring *T. cruzi* population is comprised of a highly heterogeneous pool of strains with striking intra-species variation and exhibiting wide-ranging biological characteristics in experimental settings such as distinct morphology, growth rate, curves of parasitemia, virulence, sensitivity to drugs, antigenic profile, metacyclogenesis and tissue tropism. This variability may explain the wide-ranging pathogenesis observed in *T. cruzi* infections (reviewed by Buscaglia and DiNoia [5]). Those studies recently converged on a classification that proposes the existence of six major groups in *T. cruzi*, also known as discrete typing units (DTUs): I to VI [6]. Although clonal evolution is predominant, evidence indicates that genetic exchange between parasites has occurred. CL Brener, the cloned reference strain used for the *T. cruzi* genome project, is among the strains that are the products of hybridization events. More recently, genome sequences from several other *T. cruzi* strains have been reported [7,8], allowing comparative analyses that are essential for the understanding of the variable courses of infection and clinical forms of Chagas disease. Most importantly, the study of *T. cruzi* genetic diversity provides essential information for the identification of new drug targets and antigenic candidates for better diagnostic tools and vaccine development.

In addition to investigating genomic variation, differences in gene expression also provide valuable information necessary to fully understanding the complex patterns of interaction between different *T. cruzi* strains and their hosts. Being part of a group of early branching eukaryotes, *T. cruzi* possesses unique mechanisms to control the expression of its repertoire of about 12,000 genes. Similar to other trypanosomatid genomes, the *T. cruzi* genome is constitutively transcribed into long polycistronic transcripts containing sequences from several functionally unrelated genes. Mature, monocistronic mRNAs for protein-coding genes are produced after the pre-mRNA is processed through trans-splicing and polyadenylation, two coupled reactions determined by polypyrimidine-rich tracts located within intergenic regions (see Araújo and Teixeira, 2011 for a review [9]). As a consequence of polycistronic transcription and the scarcity of typical RNA polymerase II promoters, *T. cruzi* gene expression is regulated at the post-transcriptional level. Various studies have characterized regulatory sequences present in the 3' untranslated regions (UTRs) of different genes, which act as protein-binding sites and, together with RNA binding proteins, are key elements in the control of individual mRNA steady levels during the parasite life cycle [9].

Genome-wide expression studies showing global changes in gene expression during the *T. cruzi* life cycle not only help to unravel the processes involved in the parasite's interaction with different hosts but also provide an invaluable tool for the understanding of molecular mechanisms controlling gene expression in a group of early branching eukaryotes. Steady-state transcriptome analyses, based on microarray hybridizations of cDNAs derived from epimastigotes, metacyclic trypomastigote, tissue culture derived trypomastigotes and axenically-induced amastigotes of the Brazil strain show that almost 5000 genes exhibit statistically significant up or down regulation in at least one of the four life cycle stages [10]. More recently, RNA-seq analysis of the Y strain during infection of primary human foreskin fibroblasts revealed that ~15 to 30% of the predicted protein-coding genes undergo significant changes in gene expression when epimastigotes, trypomastigotes and amastigotes were compared. Focusing on the parasite energy metabolism, this study highlighted the importance of mitochondrial electron transport for intracellular survival of *T. cruzi* amastigotes [11]. Similar RNA-seq analysis of trypomastigotes, amastigotes and epimastigotes of the Dm28c strain showed an increased expression of genes encoding surface proteins in extracellular trypomastigotes [12]. In a comparative transcriptome analysis of *T. cruzi* strains belonging to the two main genetic groups present in the *T. cruzi* population, the Sylvio (DTU I) and Y (DTU II) strains, Houston-Ludlam et al.

(2016) showed the existence of striking similarities among the differentially expressed core genes [13]. However, genome-wide analyses comparing *T. cruzi* strains with clear differences in their virulence properties have yet to been conducted.

Here, we present a comparative transcriptomic analysis, across multiple life cycle stages for two nearly isogenic *T. cruzi* cloned strains, CL Brener and CL-14, which were originally cloned from the parental CL strain [14]. In contrast to CL Brener, CL-14 is non-infectious and non-pathogenic *in vivo*, even when inoculated into newborn [15,16] or immune deficient, $CD8^{-/-}$ mice [17] that are otherwise highly susceptible to *T. cruzi* infection. Importantly, after inoculation in adult animals, CL-14 prevents the development of parasitemia and mortality following challenge with the virulent CL strain [16,18,19]. Since vaccination with live CL-14 induces a potent and long-lasting parasite-specific humoral and T-cell mediated immunity against challenge with highly virulent strains of *T. cruzi*, the immunological adjuvant properties of the CL-14 clone is being explored as a possible vaccine vector for induction of T-cell mediated immunity against other diseases [17]. Preliminary comparisons of the CL Brener and CL-14 genomes, based on shotgun sequences, revealed only few differences in the core gene content of both strains. Although a direct comparative study between the two fully assembled genomes is still under way, our initial analyses showed that the CL-14 genome, which is also a hybrid, TcVI genome, is 99.73% identical to CL Brener (S. Teixeira, in preparation). Thus, in the absence of major differences identified so far at the genome level, we analyzed the CL-14 and CL Brener transcriptomes and compared them with the aim of identifying changes in gene expression that could explain the differences in virulence between the two strains.

Results and discussion

Distinct outcomes of the *in vitro* infection with CL Brener and CL-14 cloned strains

The availability of two genetically similar *T. cruzi* strains that display marked differences in virulence presents a unique opportunity to identify fundamental phenotypic differences associated with the colonization and completion of the intracellular infection cycle in mammalian host cells (Fig 1A). To characterize the dynamics of intracellular infection by *T. cruzi* CL Brener and CL-14 cloned strains, parallel infections of low passage human fibroblasts (HFF) were established and monitored for several days post-infection. The capacities of CL Brener and CL-14 trypomastigotes to infect HFF were comparable with a similar fraction of the host cell monolayer colonized by these parasites (Fig 1B). In contrast, analysis of the number of intracellular amastigotes over time revealed significant differences in amastigote growth rates, with CL-14 lagging behind the faster growing CL Brener (Fig 1C). The slower growth of CL-14 amastigotes correlated with delayed release of trypomastigotes from infected cells upon completion of the intracellular growth and differentiation cycle (Fig 1D). As a consequence, significantly fewer trypomastigotes were released into the supernatant of CL-14-infected cells as compared to monolayers infected with CL Brener (Fig 1D). These findings suggest that the intracellular stages of CL-14 exhibit both growth and developmental defects, which may underlie the inability of CL-14 parasites to establish patent infection *in vivo*.

Global transcriptome assessment from developmental stages in both strains

To further characterize the differences in host cell infection phenotype, a comparative transcriptome analysis was performed focused on the mammalian-infective stages of CL Brener and CL-14. Total RNA was isolated from extracellular trypomastigotes and from fibroblasts

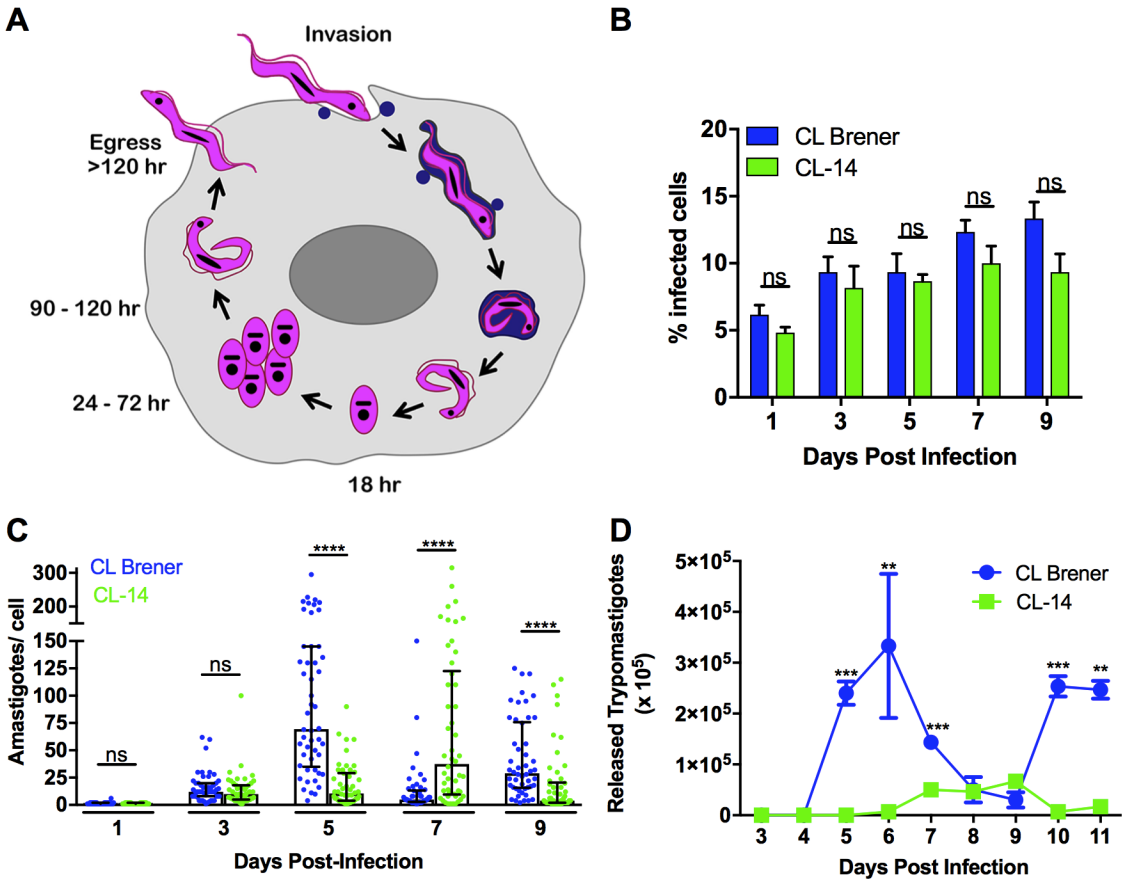


Fig 1. Comparative course of infection of human fibroblasts with *T. cruzi* CL Brener and CL-14. Intracellular *T. cruzi* life stages in mammalian cells: extracellular *T. cruzi* trypomastigotes actively penetrate mammalian cells where they differentiate into amastigotes and escape the vacuole before beginning to proliferate at ~24 hours post-infection **(A)**. Amastigotes replicate intracellularly in the host cell cytoplasm for 3–5 days, and then differentiate into motile trypomastigotes that are eventually released upon disruption of the host cell. Tissue-culture derived trypomastigotes of the CLB and CL-14 *T. cruzi* strains were similarly able to establish intracellular infection in cultured HFF **(B)** but exhibited markedly different intracellular growth dynamics as amastigotes. **(C)** Differences in the peak day of trypomastigote release from infected monolayers **(D)**.

https://doi.org/10.1371/journal.ppat.1006767.g001

infected with each strain in parallel, at time points post-infection that capture active amastigote replication (60 hr) and the transition from amastigotes to trypomastigotes (96 hr) (Fig 1A). The mRNA populations derived from biological triplicates of each sample were sequenced on the Illumina platform and the total number of reads generated from all libraries as well as the percentages of reads mapped to the *T. cruzi* CL Brener reference genome are shown in S1 Table. Consistent with the general trend observed in the numbers of intracellular amastigotes per cell, the proportion of parasite reads mapping to the reference genome at 96 hours post-infection (hpi) was smaller in cells infected with CL-14 (12.1% on average) than with CL Brener (37.0% on average). As shown in Fig 1C, the numbers of amastigotes per cell at day 5 post infection were higher in CL Brener infected cells than in CL-14 infected cells. Also consistent with data shown in Fig 1C, a smaller difference (6.7% versus 10.7%, respectively) was observed at 60 hpi, when the numbers of intracellular amastigotes were similar between the two infections.

Mapped sequencing data derived from all libraries were analyzed using a variety of methods, including principal component analysis (PCA) and hierarchical clustering, to inspect the

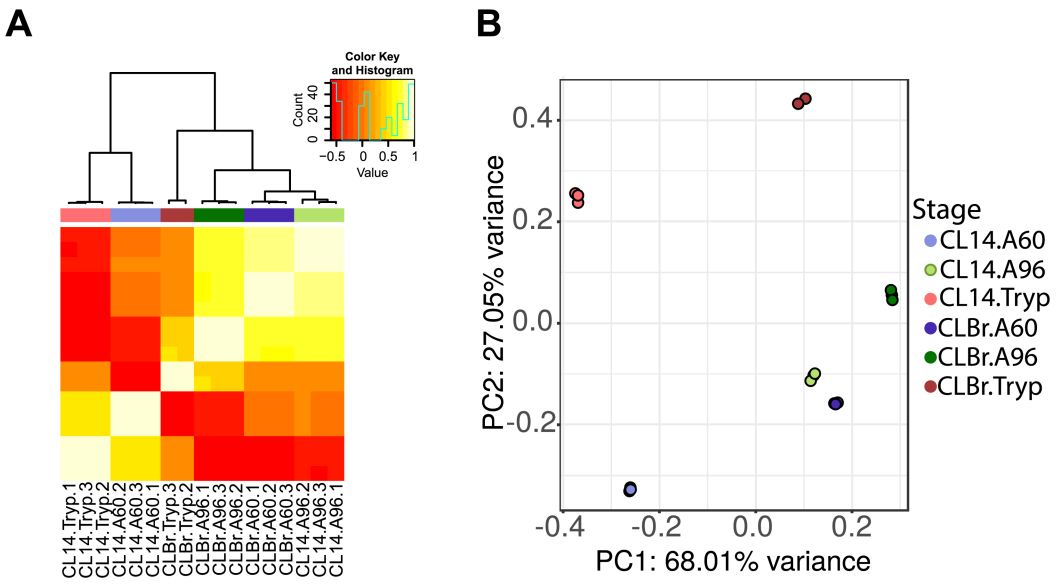


Fig 2. Global statistical assessment of biological replicates. Heat-map **(A)** and Principal Component Analysis (PCA) plots **(B)** of RNA-Seq data generated from the libraries mapped to the *T. cruzi* genome following removal of rRNA/tRNA features. Once the outlier sample was removed and data passed through normalization and surrogate variable analysis, the strong clustering by condition became evident in both analyses. In both plots, each sample is color coded by developmental stage/strain (Tryp: trypomastigotes; A60: amastigotes collected 60 hpi; A96: amastigotes collected 96 hpi).

https://doi.org/10.1371/journal.ppat.1006767.g002

relationships between samples and to identify potential outliers. The resulting PCA plot and heat maps (S1 Fig), after normalization, showed the expected clustering between biological triplicates with the exception of one sample generated from RNA isolated from CL Brener trypomastigotes (CLB.Tryp.1), which was identified as an outlier and excluded from subsequent analyses. Following normalization and surrogate variable analysis, the hierarchical clustering of the remaining samples revealed an intriguing set of observations (Fig 2A). The global transcriptome profiles of the CL-14 amastigotes at 96 hpi, while resembling their CL Brener counterparts, clustered tightly with the profiles of CL Brener amastigotes at 60 hpi. Additionally, the profiles from CL-14 trypomastigotes displayed a greater similarity to mammalian intracellular amastigote stages (from both strains) than the CL Brener trypomastigotes. The PCA plot displayed similar characteristics (Fig 2B). These findings suggested that a delayed expression of a large number of genes might be occurring in CL-14 amastigotes.

Temporal patterns of surface proteins expression may play a role in parasite virulence

The numbers of *T. cruzi* genes that emerged as differentially expressed (DE) with adjusted P value of <0.05, in pairwise comparisons of similar developmental stages of CL Brener and CL-14, are shown in Fig 3 (see S2, S3 and S4 Tables for the complete datasets). For both parasite strains, comparisons between an intracellular stage (A60 or A96) and the extracellular trypomastigote stage (Tryp) yielded a larger set of differentially expressed (DE) genes than similar contrasts between the intracellular stages (A60 vs A96) (Fig 3). This finding, which agrees with our recent characterization of global transcriptome changes in different stages of the *T. cruzi* Y strain parasites [11], was anticipated considering the extensive morphological and biochemical differences that exist between the intracellular and extracellular parasite life stages.

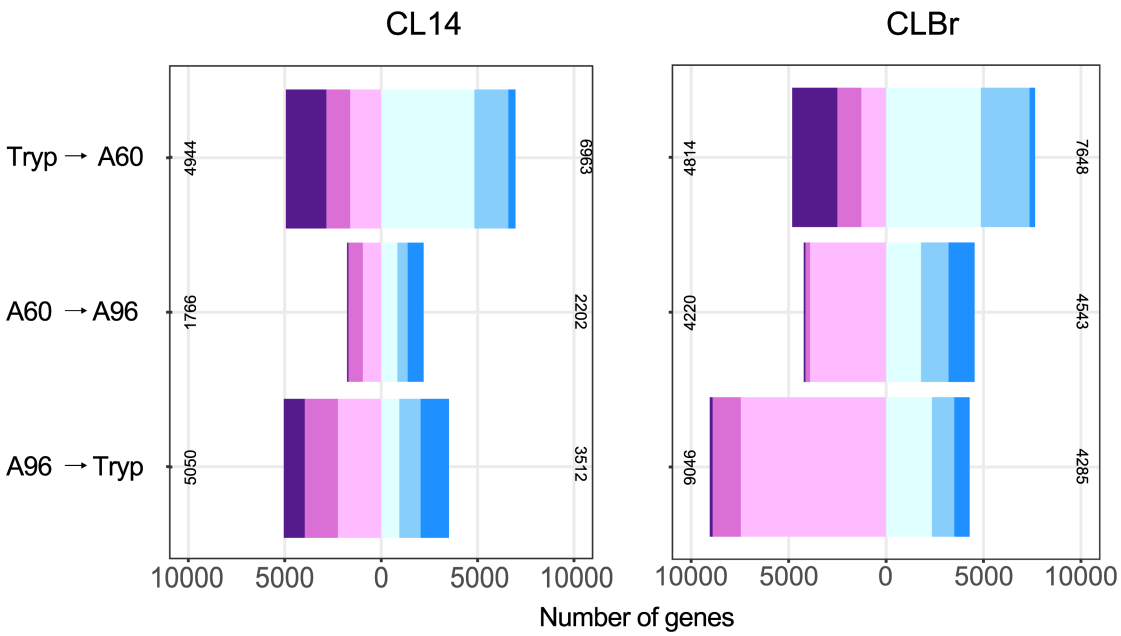


Fig 3. Differential gene expression across *T. cruzi* CL-14 and CL Brener developmental stages. Bar plots of the numbers of genes deemed significantly differentially expressed (adjusted P value <0.05) in (A) CL-14 and (B) CL Brener. The numbers of genes in each category are defined as: log2 fold changes |0–1| (light cyan for positive, light plum for negative), |1–2| (light sky blue for positive, orchid for negative), and |2+| (dodger blue and purple respectively). The diverging patterns of expression are shown by changing bar patterns.

https://doi.org/10.1371/journal.ppat.1006767.g003

As demonstrated (Fig 1B–1D), the CL Brener strain completes its intracellular life cycle in mammalian host cells more rapidly and more efficiently than CL-14: similar differences are predicted to be reflected in the global gene expression patterns. Changes in the CL Brener global transcriptome appeared to be more pronounced with twice the number of significantly up- and down-regulated genes when compared to CL-14 (Fig 3), particularly when amastigotes at 60 hpi and 96 hpi were compared. The most significant differences appeared to lie in the temporal patterns of expression of some multi-gene family encoding surface proteins, which constituted 26% to 89% of highly DE genes (>2 fold). This is best illustrated when members of multigene families of surface proteins such as trans-sialidase (TS), mucin, mucin-associated surface proteins (MASP) and GP63 were color-coded on MA plots for all three DE contrasts (trypomastigote vs. 60 hpi amastigotes, 60 hpi vs. 96 hpi and 96 hpi amastigotes vs. trypomastigotes) (Fig 4). In the contrast representing the transition of extracellular trypomastigotes to amastigotes at 60 hpi (Tryp to A60), an expected down-regulation of genes encoding TS, MASP and mucin was conspicuous for both strains and there were few differences in the behavior of most gene families in CL-14 when compared to CL Brener. The latter observation was supported by a two-sample Kolmogorov-Smirnov (K-S) test, which quantifies and tests the distance between two distributions and a Welch's test of the mean log fold changes (logFC) among the genes in each family (S5 Table). The resulting K-S distance values were small for MASP, mucin, RHS and GP63 and TS (D = 0.05, 0.05, 0.7, 0.09, respectively) and so were the differences in the mean logFC values. A clear and diverging pattern emerged when the selected multi-gene families were compared during differentiation from 60 to 96 hpi amastigotes. Genes encoding MASP, mucin and TS were up-regulated *en masse* as CL Brener began its transformation back to trypomastigotes while showing little change in CL-14 (D = 0.39, 0.35 and 0.27, respectively). A careful examination of the mean logFC change values (S5 Table)

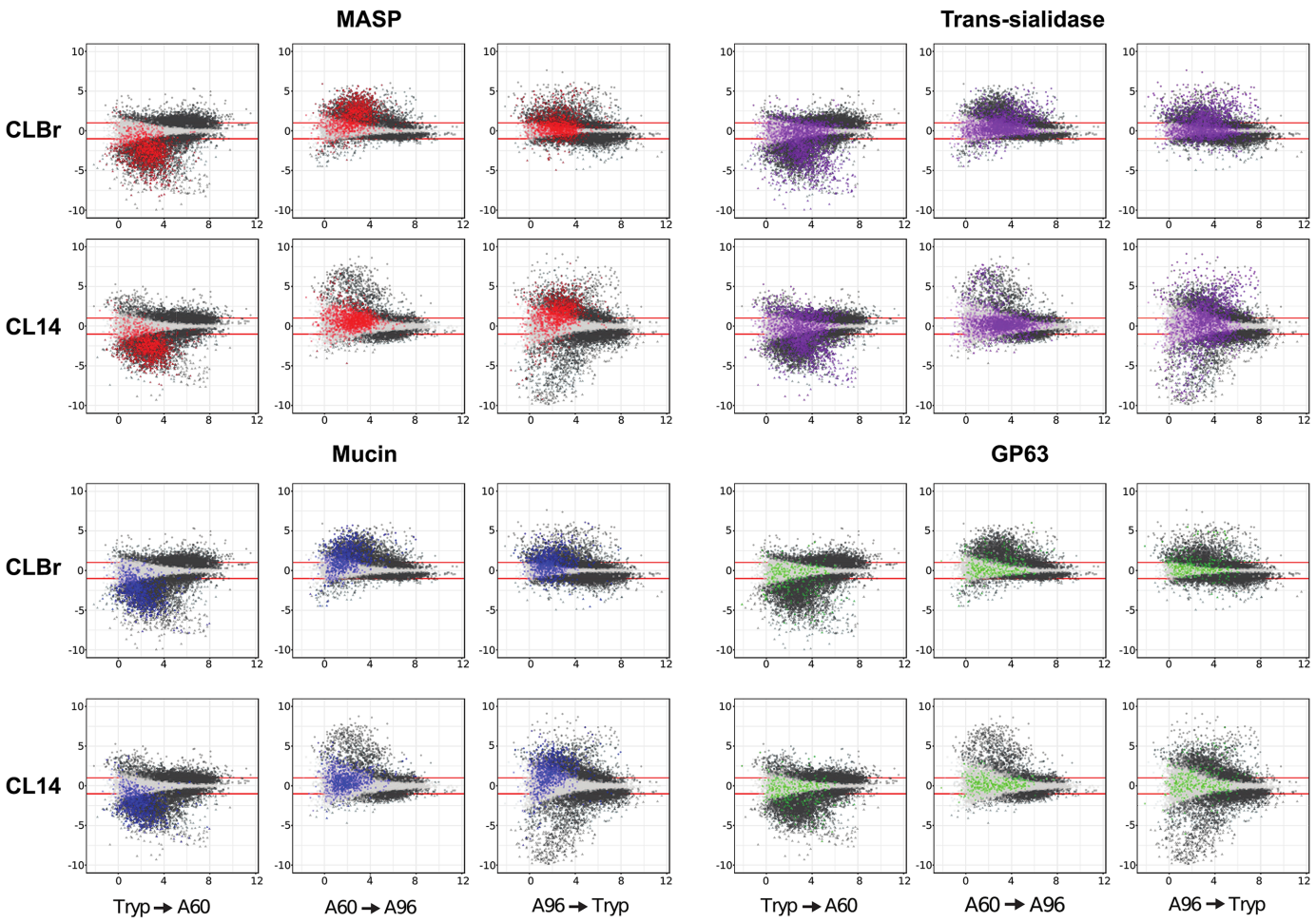


Fig 4. Comparative global transcriptional expression patterns of *T. cruzi* genes encoding polymorphic cell surface proteins in CL Brener and CL-14. MA plots depicting the log2 fold change (logFC) of genes against the average expression level during the transition of CL Brener and CL-14 across developmental stages. Each dot represents one gene and colored dots represent members of the four of the six largest *T. cruzi* gene families analyzed: MASP (red), Mucin (blue), Trans-sialidase (purple), and GP63 (green). Dots above and below the red lines represent differentially expressed genes (logFC >1).

https://doi.org/10.1371/journal.ppat.1006767.g004

reveals a more vigorous up-regulation of MASP and mucin genes when compared to TS genes, which may reflect the roles played by these cell surface proteins during parasite development in the mammalian host. The most striking observation was a delayed up-regulation of CL-14 genes encoding MASP, mucin and TS, marking a slow transition of amastigotes to trypomastigotes when compared to CL Brener and suggesting that the processes that regulate the expression of this specific group of genes encoding parasite surface proteins may be a major factor underlying the avirulent phenotype of CL-14. It is also possible that a smaller proportion of CL-14 amastigotes differentiates into trypomastigotes, resulting in the release of fewer trypomastigotes in the extracellular milieu. Members of the DGF family displayed a similar pattern of expression in CL Brener and CL-14, however their down-regulation was considerably more pronounced as CL-14 amastigotes transform into trypomastigotes (S5 Table). It is also important to note that the clear temporal asynchronicity in expression patterns of MASP, mucin and TS genes in CL-14 when contrasted to CL Brener were not observed for members of other multigene families such as GP63 and RHS.

Gene ontology enrichment analyses

A gene ontology (GO) enrichment analysis of DE genes in CL-14 and CL Brener revealed some trends similar to those observed in the transciptome analysis of *T*. *cruzi* Y strain [11]. Because the differentiation from trypomastigotes into intracellular amastigotes is accompanied by a massive down regulation of the expression of gene families encoding trans-sialidase, MASP and mucin surface proteins, we masked all DE genes that belong to any of the six largest multigenic families (MASP, mucin, DGF-1, Trans-sialidase, RHS and GP63) present in the CL Brener genome before performing the GO enrichment analysis. The transformation from extracellular trypomastigotes to intracellular replicative amastigotes was accompanied by enrichment of several biological processes including translation (GO:0006412), protein folding (GO:0006457), chromosome organization (GO:0051276), ion transport (GO:0006811), lipid biosynthesis (GO:0008610), and oxidoreductase activity (GO:0016491) in both parasite strains (Table 1). As discussed by Li et al. [11], these changes reflect the rapid proteome and membrane lipid remodeling as well as metabolic adaptations to distinct carbon sources that are required for parasite replication and intracellular survival. Interestingly, GO enrichment analyses described in the Y strain revealed that genes related to ion transport are among the most highly differentially expressed genes in amastigotes at 48 and 72 hpi compared to trypomastigotes. One of those genes (TcCLB.509197.39) belongs to the ZIP family of metal transporters that was also found highly expressed in 60 hpi amastigotes of CL Brener and CL-14 compared to trypomastigotes (logFC of 4.9 and 4.3, respectively). It is important to note that while both strains shared most of the same enriched GO processes, the numbers of genes that were significantly up regulated (with logFC>1) within the GO categories were consistently higher in CL Brener amastigotes than in CL-14. For example, whereas 109 genes related to oxidoreductase activity were up-regulated during the transformation of CL Brener trypomastigotes into 60 hpi amastigotes, only 58 genes were up-regulated in 60 hpi amastigotes of CL-14. Likewise, the numbers of genes involved in ion transport and lipid biosynthesis that were up-regulated in 60 hpi amastigotes were in general 50% lower in CL-14 when compared to CL Brener. These data indicate that the differentiation of non-dividing trypomastigotes into replicative and more metabolic active amastigotes may be occurring at a slower pace in CL-14 compared to CL Brener. On the other hand, when amastigotes at 60 hpi and 96 hpi are compared, the number of genes in the two GO categories that are upregulated in 96 hpi amastigotes of CL-14 are greater than in CL Brener, which also indicates a slower reprogramming of gene expression during intracellular differentiation of CL-14 compared to CL Brener. Among the two GO categories that were enriched with down-regulated genes when trypomastigotes from either strain differentiated into amastigotes, was a hexose transport biological process. In the current *T*. *cruzi* genome database only one gene encoding a glucose transporter (TcCLB.506355.10) is annotated, but after a more careful inpection, we identified two additional hexose transporter orthologues. Transcript levels of all three hexoxe transporters were significantly down-regulated in amastigotes from both strains (for TcCLB.506355.10, with a logFC -2.418 in CL Brener and -2.747 in CL-14), as expected, since carbohydrates do not constitute a significant energy source for intracellular amastigotes [20].

When we contrasted the GO enrichment analyses for the amastigotes' transition from 60 hpi to 96 hpi in both strains, a few notable differences were observed, including a category involved with cilium or flagellum-dependent cell motility (GO:0001539), which appeared up-regulated only in CL Brener at 96 hpi. Ciliary and flagellar motility also emerged as an enriched GO terms associated with mature *T*. *cruzi* amastigotes as early as at 48 hpi in Y strain [11]. Another interesting difference from CL Brener included an enrichment of CL-14 genes involved in chromosome organization (GO:0051276) and ion transport (GO:0006811) that

Table 1. Gene ontology (GO) categories (biological process, BP and metabolic function, MF) enriched across the *T. cruzi* CL Brener and CL-14 life cycles.

	GO Term (CL Brener)	*P* value	Num. DE	Total Num.
Trypomastigote to Amastigote 60 hpi, up-regulated				
GO:0006412	translation (BP)	2.52E-55	177	475
GO:0016491	oxidoreductase activity (MF)	6.13E-17	109	455
GO:0006811	ion transport (BP)	3.57E-16	56	165
GO:0006457	protein folding (BP)	1.87E-15	45	118
GO:0051276	chromosome organization (BP)	2.47E-12	30	70
GO:0008610	lipid biosynthetic process (BP)	5.87E-10	40	134
Trypomastigote to Amastigote 60 hpi, down-regulated				
GO:0004672	protein kinase activity (MF)	2.78E-04	32	418
GO:0008645	hexose transport (BP)	4.02E-02	1	1
Amastigote 60 hpi to Amastigote 96 hpi, up-regulated				
GO:0006565	L-serine catabolic process (BP)	3.51E-05	3	3
GO:0001539	cilium or flagellum-dependent cell motility (BP)	7.82E-03	3	12
Amastigote 60 hpi to Amastigote 96 hpi, down-regulated				
GO:0006457	protein folding (BP)	1.01E-05	8	118
GO:0006730	one-carbon metabolic process (BP)	2.30E-04	2	3
GO:0006811	ion transport (BP)	1.56E-02	5	165
GO:0016491	oxidoreductase activity (MF)	2.00E-02	9	455
Amastigote 96 hpi to Trypomastigote, up-regulated				
GO:0004672	protein kinase activity (MF)	2.13E-04	18	418
GO:0008645	hexose transport (BP)	1.67E-02	1	1
Amastigote 96 hpi to Trypomastigote, down-regulated				
GO:0016491	oxidoreductase activity (MF)	1.80E-15	69	455
GO:0006457	protein folding (BP)	1.97E-12	29	118
GO:0006757	ATP generation from ADP (BP)	1.41E-10	14	31
GO:0008610	lipid biosynthetic process (BP)	3.25E-10	28	134
GO:0006811	ion transport (BP)	2.71E-09	30	165
GO:0051276	chromosome organization (BP)	1.56E-08	18	70
GO:0009116	nucleoside metabolic process (BP)	9.22E-05	18	124
GO:0006412	translation (BP)	1.49E-03	41	475
	GO Term (CL-14)	***P* value**	**Num. DE**	**Total Num.**
Trypomastigote to Amastigote 60 hpi, up-regulated				
GO:0006412	translation (BP)	1.39E-73	166	475
GO:0051276	chromosome organization (BP)	3.56E-13	26	70
GO:0006811	ion transport (BP)	5.08E-07	31	165
GO:0006457	protein folding (BP)	5.11E-07	25	118
GO:0016491	oxidoreductase activity (MF)	9.90E-06	58	455
GO:0008610	lipid biosynthetic process (BP)	2.99E-03	19	134
Trypomastigote to Amastigote 60 hpi, down-regulated				
GO:0004672	protein kinase activity (MF)	7.48E-04	29	418
GO:0008645	hexose transport (BP)	3.68E-02	1	1
Amastigote 60hpi to Amastigote 96 hpi, up-regulated				
GO:0051276	chromosome organization (BP)	2.79E-04	10	70
GO:0006811	ion transport (BP)	7.65E-03	14	165
Amastigote 60hpi to Amastigote 96 hpi, down-regulated				
GO:0016491	oxidoreductase activity (MF)	1.40E-02	24	455

(*Continued*)

Table 1. (*Continued*)

	GO Term (CL Brener)	*P* value	Num. DE	Total Num.
Amastigote 96 hpi to Trypomastigote, up-regulated				
GO:0004672	protein kinase activity (MF)	1.40E-02	30	418
GO:0008645	hexose transport (BP)	4.86E-02	1	1
Amastigote 96 hpi to Trypomastigote, down-regulated				
GO:0006412	translation	1.73E-47	151	475
GO:0051276	chromosome organization	8.50E-15	31	70
GO:0006811	ion transport	1.05E-11	46	165
GO:0006720	isoprenoid metabolic process	3.47E-06	11	23
GO:0016491	oxidoreductase activity	2.60E-05	71	455
GO:0006457	protein folding	1.62E-04	24	118
GO:0008610	lipid biosynthetic process	1.81E-04	27	134
GO:0006730	one-carbon metabolic process	1.00E-03	3	3
GO:0009116	nucleoside metabolic process	1.30E-03	23	124

(P-value cutoff of <0.05; Number of DE genes and total number of genes in each GO term are shown)

https://doi.org/10.1371/journal.ppat.1006767.t001

was sustained in CL-14 amastigotes at 96 hpi. We noted earlier that genes encoding several cell surface proteins were up-regulated *en masse* as CL Brener began its transformation back to trypomastigotes, i.e., when 60 hpi and 96 hpi amastigotes were compared. In CL-14, the induced expression of these surface gene families became evident only when 96 hpi amastigotes and trypomastigotes were compared. The GO enrichment analyses indicated that in addition to genes encoding cell surface proteins, amastigotes of CL-14 at 96 hpi did not express genes required for the last steps of amastigote-to-trypomastigote differentiation, such as genes involved with flagellum-dependent motility (TcCLB.510955.40, TcCLB.509005.10, TcCLB.508793.14). Also, genes related to protein folding (GO:0006457) and ion transport (GO:0006811), which were down-regulated in 96 hpi amastigotes compared to 60 hpi amastigotes in CL Brener did not show significant differences in gene expression when 60 hpi and 96 hpi amastigotes of CL-14 were compared.

When GO enrichment was examined for the differential expression from amastigotes at 96 hpi to trypomastigotes, an expected enrichment of protein kinase activity (GO:0004672) and hexose transport (GO:0008645) was identified amongst the upregulated genes in both strains. These finding are in agreement with data that showed that glucose is the main energy source for the parasite in the bloodstream and also with increased protein modification, such as phosphorylation of MASPs, that is required for the abundant expression of cell surface glycoproteins in trypomastigotes [12,20,21,22]. Also consistent with a decreased metabolism and macromolecular biosynthesis of the non-replicative trypomastigotes, down-regulated genes in both strains were enriched in protein translation (GO:0006412), protein folding (GO:0006457), ion transport (GO:0006811), oxidoreductase activity (GO:0016491), lipid biosynthesis (GO:0008610) and chromosome organization (GO:0051276). Hence, in addition to confirming the associations between gene expression and the metabolic differences that are known to occur during the intracellular differentiation, our GO enrichment analyses revealed further differences in gene expression between CL Brener and CL-14 during the transition from early to late stages of the intracellular amastigote development. It can be assumed that, in addition to a distinct program controlling the expression of multigene families encoding cell surface glycoproteins, other differences in gene expression related to changes in the parasite metabolism

and cellular structures may also contribute to the distinct behavior of the two strains during infection of mammalian cells.

Differential expression of RNA-binding proteins

To further investigate the regulatory mechanisms controlling the distinct transcriptional programs that drive the development from intracellular amastigotes to the extracellular trypomastigote stage in the two cloned strains, we analyzed the expression of RNA-binding proteins (RBPs) that are differentially expressed throughout the infection of the host cell by CL Brener and CL-14. Using homology searches to four protein domains known to be present in RNA binding proteins (PUF (Pumilio), zinc finger, RNA recognition motif (RRM) and Alba domains), we identified a total of 253 *T*. *cruzi* sequences representing 147 genes (since two alleles were identified for most genes), which are shown in S6 Table. This list includes all genes encoding *T*. *cruzi* RBPs containing the RRM motif identified by Gaudenzi et al. (2005) [23] and Guerra-Slompo et al. (2012) [24], Pumilio-type RNA binding proteins identified by Caro et al. (2006) [25] and Zinc Finger containing proteins identified by Kramer et al. (2010) [26], with an additional 13 genes. Among all the putative RBP-encoding genes identified, six genes were differentially expressed in CL Brener (logFC>2, P value < 0.05) during differentiation between 60 hpi and 96 hpi amastigotes and from amastigotes to trypomastigotes (Fig 5A–5F). Amongst them is the gene encoding a CCCH zinc finger (ZFN) motif-containing protein (TcCLB.511511.6) that was significantly upregulated in trypomastigotes from CL Brener as well as in CL-14 (Fig 5A). Previous studies with the Dm28c strain of *T*. *cruzi* indicate that this RBP acts as a regulatory factor possibly involved with metacyclogenesis [27], whereas the *T*. *brucei* homologue TbZFN1 (Tb927.6.3490) was identified as a factor that is upregulated during differentiation from bloodstream to procyclic forms [28].

Five transcripts encoding RBPs are upregulated in amastigotes from both strains, indicating that they may control expression of genes involved with replication and intracellular survival of the parasite in the mammalian host cell. TcCLB.511727.190 encodes a RRM containing protein, which was significantly upregulated in 60 hpi and 96 hpi amastigotes compared to trypomastigotes of CL Brener and CL-14 (Fig 5B). Its *T*. *brucei* homologue DRBD18 (Double RNA Binding Domain, encoded by Tb927.11.14090) can be modified by arginine methylation [29]. TcCLB.469785.40, identified as amastigote-specific not only in CL Brener and CL-14 but also in the Y strain [11], is also a member of the DRBD family whose *T*. *brucei* ortholog (Tb927.6.3480) has been identified as the post-transcriptional repressor DRBD5 [30] (Fig 5C). Similar to TcCLB.469785.40, transcript levels of the RBP encoded by TcCLB.509551.60 is increased in 60 hpi amastigotes of CL Brener compared to 96 hpi amastigotes and trypomastigotes as well as in amastigotes of the Y strain up to 12 hpi [11] (Fig 5D).

TcCLB.504005.6, annotated as RBP5 (Fig 5E) encodes an RBP that is upregulated in 96 hpi amastigotes of CL Brener and CL-14, as well as in the Y strain amastigotes at 72 hpi amastigotes but not at 48 hpi [11]. Similarly, transcript levels of the *T*. *cruzi* PUF9 (TcCLB.506563.10) are also up-regulated in CL Brener 96 hpi amastigotes as well as in Y strain amastigotes at 72 hpi compared to amastigotes at early stages of the infection (Fig 5F). In *T*. *brucei*, PUF9 (Tb927.1.2600) stabilizes a group of transcripts during S-phase and may be involved with the temporal coordination of nuclear and kinetoplast replication [31]. It is therefore tempting to speculate that the RBPs encoded by TcCLB.504005.6 and TcCLB.506563.10 act as regulatory factors involved with the final steps of amastigote-trypomastigote differentiation and the release of trypomastigotes from infected cells.

Because of their critical role in regulating gene expression in trypanosomatids, we searched for RBPs whose expression significantly differs in the CL Brener and CL-14 transcriptomes. In

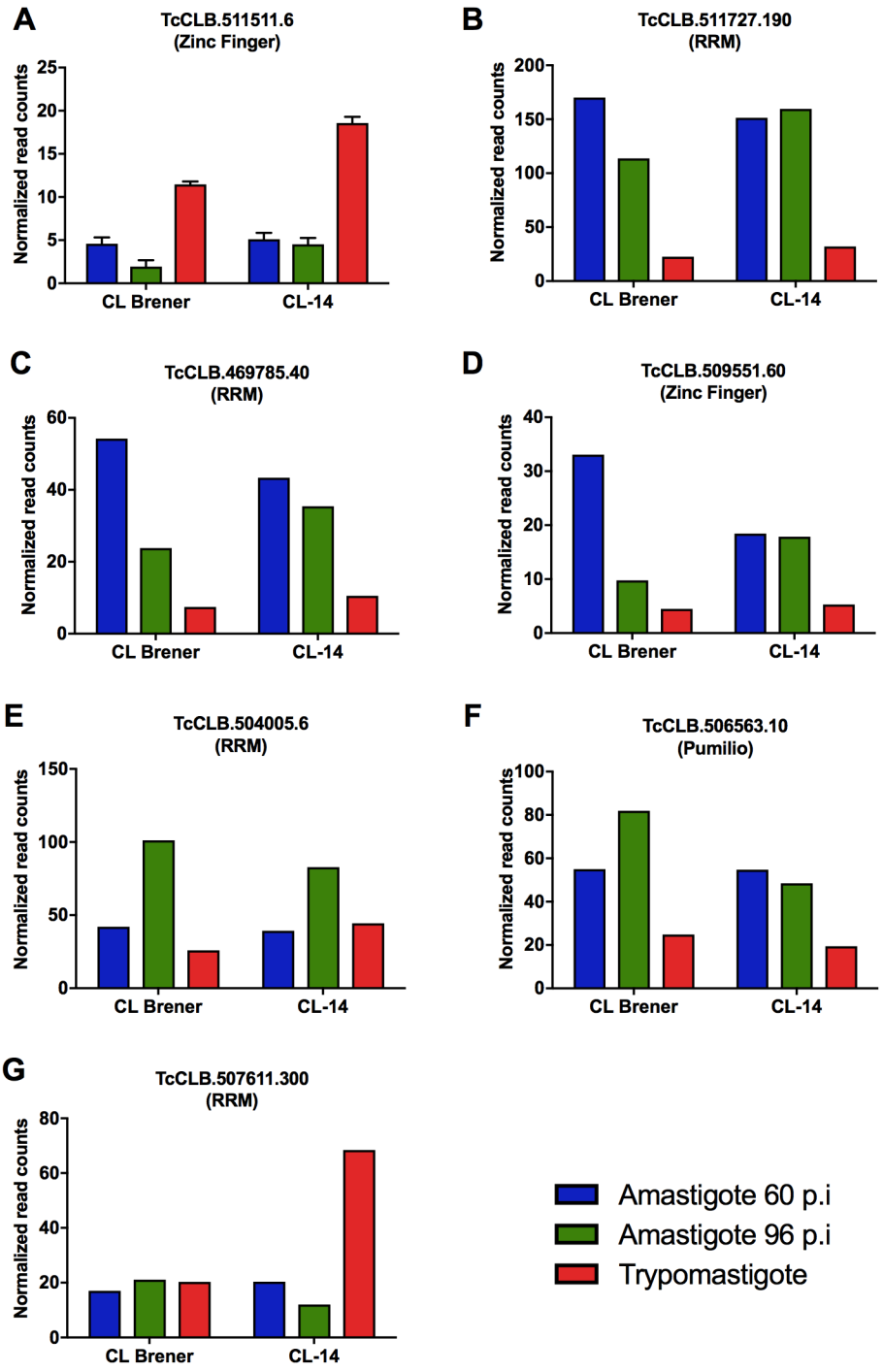


Fig 5. Expression patterns of selected genes encoding RNA binding proteins. Expression values for genes encoding RPBs were extracted from S4 Table. The selected genes were observed to be up-regulated in CL Brener and CL-14 trypomastigotes **(A)**, in CL Brener and CL-14 amastigotes **(B-F)**, and only in trypomastigotes from CL-14 **(G)**. Only genes encoding proteins with RNA binding domains whose transcripts present logFC > 2 comparing amastigotes at 60 hpi and 96 hpi and trypomastigote stages are shown. When sequences from both haplotypes are present, the sequences shown in the graphs correspond to Esmeraldo-like alleles, but similar expression patterns were obtained with the non-Esmeraldo alleles.

https://doi.org/10.1371/journal.ppat.1006767.g005

contrast to all six differentially expressed RBP genes shown in Fig 5A–5F, TcCLB.507611.300 is constitutively expressed in the mammalian stages of CL Brener as well as in the Y strain [11], but its transcript levels are significantly increased in trypomastigotes of CL-14 (Fig 5G). TcCLB.507611.300 encodes a spliceosomal protein that is part of the U1A small nuclear ribonucleoprotein complex as indicated by the studies with the *T. brucei* ortholog (Tb927.10.830) [32]. In addition to further investigating whether the differences observed in the expression of CL Brener and CL-14 RBP genes, such as TcCLB.469785.40 and TcCLB.509551.60, which may be involved with the control of steady state levels of their target transcripts, further studies are also needed to verify whether the differential expression of an RBP involved in mRNA processing may cause any impact on the expression of genes related to the virulence of CL-14 trypomastigotes.

Constitutive expression of a trans-sialidase gene in CL-14 affects amastigote to trypomastigote differentiation

Because data shown in Fig 4 and S5 Table indicated that several genes encoding surface protein families are been expressed at later time points during intracellular infection of CL-14 compared to CL Brener, we sought to test whether altering the expression of a member of these multigene families might affect parasite differentiation inside the host cell. We selected the trans-sialidase (TS) gene Tc00.1047053509495.30 (whereby named TS95.30) because it belongs to TcS group I and encodes an enzymatically active TS that contains the C-terminal SAPA (Shed Acute Phase Antigen) repeats known to be expressed only in the mammalian stages of the parasite life cycle [33]. Our RNA-seq analyses confirmed the expression of TS95.30 in amastigotes and trypomastigotes of CL Brener and CL-14 but also highlighted distinct expression levels of TS95.30, which is highly upregulated (FC = 9.8) in 96 hpi amastigotes compared to 60 hpi in CL Brener, and less so in CL-14 (FC = 3.5).

The complete open reading frame of TS95.30 was cloned in the pROCKNeo vector [34] and transfected into CL-14 epimastigotes using the linearized form of the vector shown in Fig 6A. As described previously, transfection with linearized pROCKNeo vector results in the integration of the DNA construct in the tubulin locus and, because signals for mRNA processing of the foreign gene are derived from the constitutively expressed housekeeping genes *TcP2*β (at the 5' end) and *gapdh* (at the 3' end), the pROCKNeoTS transgene can be constitutively expressed in all parasite life cycle stages, including epimastigotes [34]. After selecting a neomycin-resistant population as well as transfected cloned cell lines, total RNA and protein extracts from these cultures were subjected to northern (Fig 6B) and western blot analyses (Fig 6C). For northern blot, we used a probe derived from the 3' region of the TS95.30 gene and for the western blot we used an anti-SAPA antibody that detects the expression of members belonging to TS group 1 sub-family. As shown in Fig 6B, two cloned epimastigote cell lines derived from G418-resistant, CL-14 parasites transfected with the TS gene (clones TS 14 and TS 22) showed increased levels of TS mRNA compared to untransfected CL-14 or with epimastigotes from CL Brener. These two transfected cell lines plus two additional ones (TS 16 and TS 18), all of them showing increased TS protein expression compared to WT CL-14 (Fig 6C), were selected and tested in *in vitro* infection assays. No differences in the growth curves of epimastigote cultures from all four transfected cell lines compared to WT CL-14 and CL Brener epimastigotes were observed (S2 Fig).

The infection capacity of the transfected parasites was tested *in vitro* by incubating Vero cells with equal numbers of tissue culture derived trypomastigotes from WT CL Brener, CL-14 and from the CL-14 transfected cell lines (TS 14 and TS 22) constitutively expressing TS95.30. As shown in Fig 6D and 6E, parasites from all four cell lines transfected with TS displayed

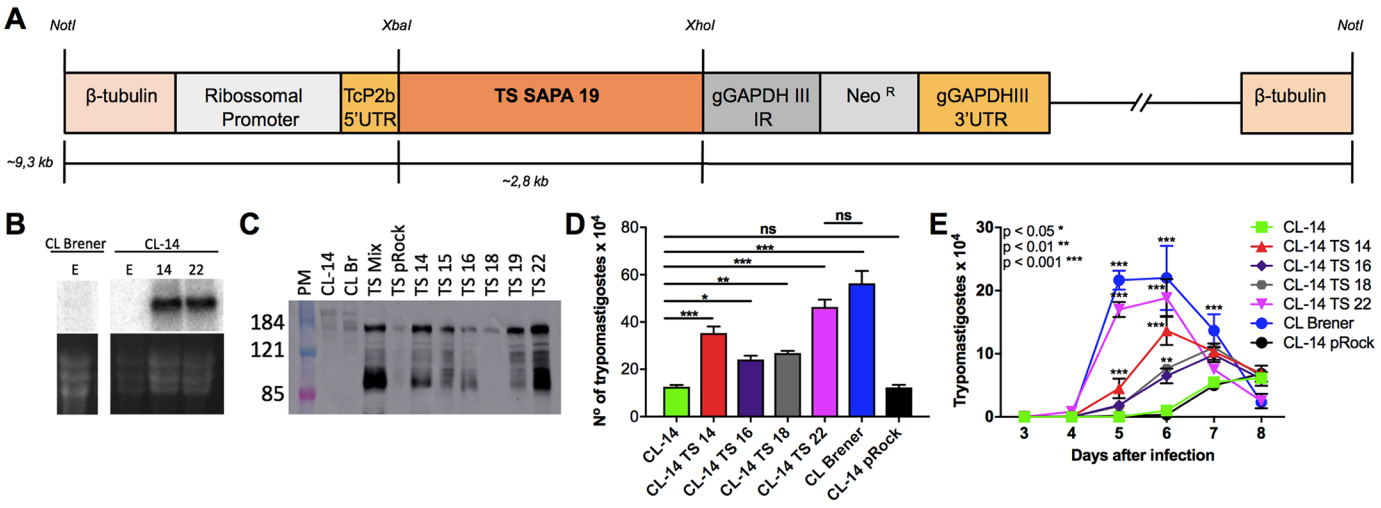


Fig 6. Constitutive expression of a TS gene in CL-14. The pROCKNeo vector used for transfection of CL-14 has the TS gene (Tc00.1047053509495.30) flanked by the ribosomal promoter and sequences containing signals for mRNA processing derived from the constitutively expressed housekeeping genes TcP2β (at the 5' end) and gapdh (at the 3' end) **(A)**. Total RNA purified from epimastigotes from WT and transgenic parasites were subjected to northern blot and hybridized with a ^{32}P-labelled probe that contains sequences corresponding to the C-terminal SAPA repeats present in the TS gene. Lower panel shows ethidium bromide staining of rRNAs in the same gel before transferring to the membrane **(B)**. Total protein extracts from epimastigotes from WT and transgenic parasites were evaluated for the expression of the transfected TS gene by western blotting with a monoclonal antibody anti-SAPA **(C)**. The infection profiles of four cloned cell lines derived from CL-14 parasites transfected with the TS gene or with the empty pROCKNeo vector, were compared to WT CL-14 and CL Brener in *in vitro* infection assays of Vero cells. Equal numbers of tissue culture derived trypomastigotes from each parasite cultures were added to Vero cell monolayers and the total number of trypomastigotes released in the supernatant **(D)** or the numbers of trypomastigotes released in the supernatant each day post-infection were evaluated over 8 days **(E)**. Five replicates for each infection experiment were performed.

https://doi.org/10.1371/journal.ppat.1006767.g006

faster kinetics of trypomastigote release in the supernatant compared to untransfected CL-14 or with CL-14 parasites transfected with the empty vector. Although no significant differences in the numbers of intracellular amastigotes were observed, the numbers of trypomastigote released by cells infected with transfected CL-14 parasites was increased compared to cells infected with WT CL-14 or with CL-14 transfected with the empty vector. This result indicates that the constitutive expression of one member of the TS gene family is sufficient to alter the differentiation from amastigote to trypomastigotes, a process that occurs intracellularly before trypomastigotes are released to infect new host cells. Although constitutive TS expression facilitates intracellular amastigote to trypomastigote differentiation and parasite egress from infected cells, constitutive expression of TS is not sufficient to restore parasite virulence. When the two cloned, transfected CL-14 cell lines TS 14 and TS 22 were inoculated into BALB/c mice, similar to WT CL-14 infected animals, no parasites were detected in the bloodstream at any time point after infection (S3 Fig). It should be noted that, cells infected with one transfected cell line, TS 22, showed a total number of trypomastigotes released in the supernatant that was statistically equivalent to the number of released trypomastigotes in CL Brener infected cells (Fig 6D). It is therefore assumed that differences regarding the gene expression profiles of trypomastigotes from both strains also contribute to determine the distinct virulence phenotypes observed during infection *in vivo*. The mechanisms by which the increased expression in CL-14 of a TS gene affects amastigote to trypomastigote differentiation need to be investigated. The work of Rubin-de-Celis and colleagues, [35] suggests that sialic acid present on lysosomal membrane glycpoproteins are targets of TS, since over-expression of TS in the Y strain facilitates escape of amastigotes from the parasitophorous vacuole. It is also possible that the constitutive expression of TS triggers other changes in gene expression or

modifications at the cell surface that affect parasite differentiation. It has been shown that glycosylphosphatidylinositol anchored proteins such as TS, mucins and MASPs are transported to the parasite surface through the contractile vacuole complex, in a trafficking process that involves the small GTPase protein Rab11 [36]. Therefore, the possibility that early overexpression of TS in CL-14 may cause changes in membrane trafficking in a way that affects membrane remodeling events required for amastigote to trypomastigote differentiation, should also be considered. These hypotheses are currently being investigated using RNA-seq analyses. As predicted by the differential expression and GO analyses, adequate expression of mucins, MASPs and/or several other genes may also be needed during late stages of intracellular development of the parasite, along the expression of TS genes, to restore CL-14 virulence in animal models of infection. Our results also suggest that a genome wide DNA manipulation strategy allowing the inspection of multiple factors will be necessary to fully address the question regarding differences in the virulence phenotype among the two *T. cruzi* strains.

Concluding remarks

Our *in vivo* infection assays confirmed previous studies, which have shown that, in contrast to an infection with the CL Brener clone, no patent parasitemia can be detected in immune competent animals that were inoculated with trypomastigotes from the attenuated CL-14 clone. However, *in vitro* infection assays showed that CL-14 trypomastigotes are able to infect and multiply within different cell types (this study tested infection of HFF and Vero cells), although the CL-14 infection resulted in lower numbers of intracellular amastigotes during early time points and a markedly reduced number of trypomastigotes released by the CL-14 infected cells compared to the CL Brener infection. One might conclude that a deficient capacity of CL-14 amastigotes to differentiate into trypomastigotes is a key factor accounting for the CL-14 attenuated phenotype. The results from the *in vitro* infection assays prompted us to investigate the molecular basis of the attenuated phenotype of CL-14 by performing gene expression analyses at two time points of the intracellular development of the two strains. As shown previously [11], once reaching the cell cytoplasm, CL Brener amastigotes replicate for about 4 days, before they differentiate back into trypomastigotes and rupture the cell. Our comparative transcriptome analyses showed that the transition of extracellular trypomastigotes to amastigotes at 60 hpi was accompanied by only few differences in gene expression between the two parasite strains, including a similar down-regulation of multi-gene families encoding TS, MASP and mucin. In contrast, large differences in the transcriptome remodeling were observed between the two strains when gene expression at 60 hpi and 96 hpi were compared, particularly regarding the expression of gene families encoding surface proteins, which are up-regulated with faster kinetics in CL Brener than in CL-14. Thus, the RNA-seq data presented in this study point to a central role of this group of proteins in the process of differentiation from the replicative amastigote to infective trypomastigotes and establishment of infection in the mammalian host.

Our RNA-seq analysis is also consistent with experimental data that shows no differences in the parasite's invasion capacity or ability to multiply in the cytoplasm of the mammalian cell. As observed during *in vivo* infection, it is possible that CL-14 amastigotes are able to multiply in a limited number of tissues of the infected animal at least during the initial stages of the infection, but the delayed expression of genes encoding surface proteins prevents the release of trypomastigotes in the bloodstream and in neighboring tissues. This is consistent with several reports showing that, although infection with CL-14 does not result in patent parasitemia, it results in efficient immunity against a lethal challenge with virulent parasites [18,19], or against a human tumor antigen if this antigen is expressed by the parasite [17].

Because CL-14 retains its capacity to multiply during the first round of infection after trypomastigote inoculation, this initial multiplication within host tissues may be sufficient to induce a highly efficient immune response that includes a balanced Th1/Th2 cytokine production that is beneficial for the host [19].

In one study addressing the basis of the non-virulent phenotype, Atayde and co-workers [37] showed that metacyclic forms of the CL strain and from the CL-14 clone express comparable levels of gp82, a member of the TS family, and gp35/50, a member of the mucin family of glycoproteins. Although we did not analyze the gene expression profile of metacyclic forms, the RNA-seq data obtained for trypomastigotes from CL Brener and CL-14 indicated that once the parasites differentiated into trypomastigotes, their gene expression profiles still showed large differences, particularly regarding genes belonging to the multi-gene families of surface proteins. In another study where the number of extracellular trypomastigotes was determined after infection of CHO-K1 cells with CL-14, Tonelli and co-workers [38] found that increased numbers of trypomastigotes are released in the supernatants if cells are kept at 33˚C instead of 37˚C. In addition, the numbers of trypomastigotes released into the medium were 2.5 to 3-fold higher if the cultures were supplemented with 20 or 200 microM proline. We confirmed the increased capacity of CL-14 to differentiate from amastigote to trypomastigote if the temperature is lowered from 37 to 33˚C (S4 Fig). The thermo-sensitive character of the differentiation process of CL-14 will be investigated using RNA-seq data collected at both temperatures. Using RNA-seq analyses, we are also planning to investigate whether the constitutive expression of one member of the trans-sialidase in CL-14 affects the expression of other genes that may also be involved with the amastigote to trypomastigote differentiation. Finally, RNA immunoprecipitation analyses using tagged versions of the three RNA binding proteins (TcCLB.507611.300, TcCLB.506563.10 and TcCLB.469785.40), identified as being differentially expressed in both strains, are also underway with the aim of identifying their target genes, which may correspond to genes directly involved with the *T*. *cruzi* differentiation program. However, the lack of a fully assembled genome poses a major limitation that needs to be resolved before we can perform a conclusive comparative analysis between these strains. Efforts towards the generation of the complete sequences of all chromosomes from CL Brener and CL-14 are underway using the long read sequencing technology by PacBio. With this new sequencing strategy we should be able to resolve all the complex regions in this parasite genome, particularly regions where large multigene families are present. Together with additional gene expression analyses, further genome studies investigating gene copy number variation will help understand the diversity of the *T*. *cruzi* population and its impact on parasite infection.

Materials and methods

Parasite cultures, RNA extraction and cDNA sequencing

The two cloned strains of *Trypanosoma cruzi*, CL Brener and CL-14 were provided by Egler Chiari, from the Federal University of Minas Gerais. Mammalian-infective stages of *T*. *cruzi* were maintained by weekly passage in monkey kidney cell line LLc-MK2 (ATCC CCL-7) grown at 37˚C and 5% CO_2 [11]. Trypomastigotes collected from infected monolayers were centrifuged for 10 min at 1,700*g* and then allowed to swim up from the pellet for 4 hrs before collecting the pure highly motile fraction in the supernatant and washing in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 2% fetal bovine serum (DMEM-2). For RNA-Seq studies, experimental infections were established in human foreskin fibroblasts (HFF; ATCC CRL-2522) by exposing sub-confluent monolayers to purified trypomastigotes (80 parasites: host cell) for 2 hours at 37˚C, 5% CO_2 followed by rinsing in PBS 3–5 times to

remove remaining extracellular parasites to achieve infection of 60–70% of host cells. Total RNA was extracted from infected HFF at 60 and 96 hours post-infection and from tissue culture derived trypomastigotes released in the supernatants of infected HFFs using the Trizol reagent (Invitrogen, CA, USA). After DNAse treatment, the RNA was further purified using the Qiagen RNeasy mini kit, poly(A)+ RNA was selected using oligo-dT following Illumina TruSeqv2 instructions and its integrity was assessed using an Agilent 2100 bioanalyzer. cDNA libraries were constructed using a TruSeq RNA Sample Prep Kit Version II (Illumina) with an average insert size of ~300 nt. Libraries were sequenced on an Illumina HiSeq 1500.

Data quality assessment and visualization

Quality assessment of 100 nucleotide paired-end reads was performed via an initial evaluation with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and biopieces (http://biopieces.org). Illumina sequencing adapters and low quality sequences were removed with Trimmomatic [39]. The remaining sequences were mapped separately against the Esmeraldo haplotype of *T. cruzi* CL Brener genome, non-Esmeraldo haplotype, and a combination of both haplotypes along with the unassigned contigs using the TriTrypDB version 9.0 genome and annotation data (http://tritrypdb.org/common/downloads/release-9.0/). Alignments were done using TopHat2 [40] version 2.0.14 and options allowing a single randomly placed mapping per read (-g 1) and without a novel splice junction search (-G gff_file). Mapping was performed with the same options for the samples collected from human cell samples using the ensembl GRCh37.62.v3 release. Alignments were compressed, sorted, and indexed using samtools [41] in order to include only proper pairs and counted against the set of annotated transcripts using HTSeq [42]. The resulting count tables were passed to R for outlier testing, sample clustering, visualization, and differential expression analyses with the hpgltools package (http://github.com/elsayed-lab/hpgltools/). A table with raw counts for all samples and conditions is available as supplementary data (S7 Table).

Differential expression and gene ontology analyses

Samples were tested for significant outliers using a mix of hierarchical clustering and principal component analysis. Non-expressed and weakly expressed genes, defined as having <2 reads per million in n of the samples, where n is the size of the smallest group of replicates (here n = 2), were removed prior to differential expression (DE) analysis. Of the 24,887 genes analyzed, 23,191 remained after applying the low-count filter. Multiple data normalization methods were tested, including a mix of: trimmed median of M-values, quantile, log transformation, and counts per million. Surrogate variables / batches were queried with a mix of methods from sva [43] and ruv [44]. Differential expression analyses were performed and compared using surrogate estimates from svaseq with limma [45], DESeq2 [46], EdgeR [47], and a statistically uninformed basic method as a diagnostic control. Genes were considered 'significant' when the limma observed |log2 fold-change| was greater than 1.0 (logFC$>$1) and the adjusted *P* value was less than 0.05. The set of significant genes for each contrast of interest was passed to Goseq [48] for the parasite samples and g:Profiler [49] for the human samples through its R interface.

Parasite transfection and characterization

Transfected CL-14 epimastigotes were obtained by cloning the complete open reading frame (ORF) of the TS gene Tc00.1047053509495.30 into the XbaI-XhoI sites of pROCKNeo plasmid [34]. *T. cruzi* epimastigotes were transfected, selected with 200 microg/mL of G418 and several G418 resistant cloned cell lines were generated as previously described [17,34]. Expression of

TS95.30 was detected by northern blot using a ^{32}P labelled PCR fragment corresponding to the C-terminal domain of the transfected gene and by western blot using an anti-TS-SAPA monoclonal antibody (kindly provided by Sergio Shenkman, Federal University of São Paulo) according to previously described protocols [17]. *In vivo* infection of BALB/c mice (purchased from Biotério Central, Federal University of Minas Gerais) were carried out as previously described [17]. *In vitro* infection of Vero cells (ATCC CCL-81) with tissue culture derived trypomastigotes of transfected parasite cell lines were done as indicated in the previous session.

Ethics statement. Animal infections were performed in strict accordance with the Brazilian laws regarding animal use (LEI No 11.794, DE 8 DE OUTUBRO DE 2008). All protocols were approved by the Committee on the Ethics of Animal Experiments of UFMG (protocol # 132/2014).

Supporting information

S1 Fig. Quality control metrics suggesting the removal of one sample. (A) Raw library sizes were plotted to check for comparable sequencing depths for all samples, (B) boxplot of the count distributions by gene. Hierarchical clustering analyses using Pearson correlation (C) and Euclidean distances (E) along with Principal Component Analyses (D) and a standard medium Pearson correlation between each sample (F) of the log2, cpm, quantile, and low-count filtered data suggested the removal of sample 'CLB.Tryp.1'.
(EPS)

S2 Fig. Growth curves of CL-14 epimastigotes transfected with a TS gene. Wild-type (WT) CL-14 epimastigotes and cloned cell lines derived from CL-14 epimastigotes transfected with the pROCK vector containing the TS gene TS95.30 (TS 14, TS 16, TS 18 and TS 22) as well as CL-14 epimastigotes transfected with the empty pROCK vector were cultivated in LIT medium for 9 days and the numbers of parasites were determined in a Neubauer chamber.
(TIFF)

S3 Fig. Infection of BALB/c mice with WT and transgenic CL-14. Groups of 5 animals were inoculated with 5 x 10^3 tissue culture trypomastigotes from WT CL-14, from transgenic CL-14 cell lines TS 14 and TS or from CL Brener trypomastigotes and parasitemia was determined during 30 days. The data shown are representative of three independent experiments.
(TIFF)

S4 Fig. Cell infection assays with CL-14 trypomastigotes at different temperatures. Vero cells (5 x 10^4) were incubated with 5 x 10^5 CL-14 trypomastigotes at 37˚ or 34˚C and 5% CO_2. After 24 hours, cells were washed with phosphate buffered saline to eliminated extracellular parasites and the numbers of trypomastigotes released in the supernatant were determined in a Neubauer chamber. Infection assays were performed in triplicates.
(TIF)

S1 Table. Summary of samples collected, mapping statistics and experimental metadata. This summarizes the experimental reference identifiers used in the experiment (columns A-C), the experimental conditions (columns D-F), experimental batches (column G), numbers of reads collected and mapped (columns H-J), and percentage mapped to the parasite and host genomes (columns K-N).
(XLSX)

S2 Table. Differential expression for all contrasts and all genes using limma. The workbook contains one worksheet for every contrast performed by limma, starting with the annotation information provided by the TriTrypDb, followed by the limma statistics for every coding-

sequence gene which passed low count filtering. A detailed legend is provided in the first worksheet.
(XLSX)

S3 Table. Lists of significant differentially expressed genes for all contrasts using limma. This workbook contains the same information as in Table S2, but only for those genes which are deemed 'significantly differentially expressed' using a |log2 fold change| > = 2.0 and adjusted p-value < = 0.05. The worksheets are named according to the contrast performed, and the direction of the fold change. The sheet 'number_changed' provides some summary information regarding the numbers of observed differentially expressed genes according to each tool employed.
(XLSX)

S4 Table. Expression values for individual genes. The metric of 'expression' was collected from limma following the application of voom for each experimental condition. It is comprised of the normalized, batch-adjusted, expression value on the log2 scale. Errors were calculated by taking the logFC from the limma identity tables and dividing by the t-statistic for each gene. The table contains annotations (columns A-E), followed by the metric of expression for each condition (columns F-K), followed by the standard error (columns L-Q) for each condition.
(XLSX)

S5 Table. Analysis of expression trends in six *T. cruzi* multigene families. A two-sample Kolmogorov–Smirnov test (K–S test) and a Welch's two sample t-test for unequal variances were run to test for significant differences in the distributions (K-S) and means (Welch's) of the logFC values for all members of each multigene family across development.
(XLSX)

S6 Table. Putative RNA-binding domain containing genes. Genes were identified using homology searches to four protein domains known to be present in RNA binding proteins: PUF (Pumilio), zinc finger, RNA recognition motif (RRM) and Alba domain. For most genes, the two alleles (Esmeraldo-like and non-Esmeraldo-like) are shown.
(XLSX)

S7 Table. Raw counts for all samples. The workbook comprises 7 sheets. The first, 'legend', contains a legend and copy of the experimental design. The 'raw_reads' sheet contains the gene annotation data along with numbers of raw counts observed for each gene. The 'raw_graphs' a series of initial plots describing the raw data. The 'norm_data' provides the counts after the normalization applied as stated in the legend (log2, quantile, sva, filtered). The 'norm_graphs' provides the same set of plots describing the normalized data. Finally, the 'median_data' provides a set of median values of the data after normalization by experimental conditions.
(XLSX)

Author Contributions

Conceptualization: Caroline Junqueira, Ricardo T. Gazzinelli, Barbara A. Burleigh, Najib M. El-Sayed, Santuza M. R. Teixeira.

Data curation: A. Trey Belew, Caroline Junqueira, Gabriela F. Rodrigues-Luiz, Bruna M. Valente, Antonio Edson R. Oliveira, Rafael B. Polidoro, Barbara A. Burleigh, Najib M. El-Sayed.

Formal analysis: A. Trey Belew, Caroline Junqueira, Gabriela F. Rodrigues-Luiz, Antonio Edson R. Oliveira, Luciana W. Zuccherato, Ricardo T. Gazzinelli, Barbara A. Burleigh, Najib M. El-Sayed, Santuza M. R. Teixeira.

Funding acquisition: Barbara A. Burleigh, Najib M. El-Sayed, Santuza M. R. Teixeira.

Investigation: Caroline Junqueira, Bruna M. Valente, Antonio Edson R. Oliveira, Rafael B. Polidoro, Daniella C. Bartholomeu, Barbara A. Burleigh, Najib M. El-Sayed, Santuza M. R. Teixeira.

Methodology: A. Trey Belew, Caroline Junqueira, Gabriela F. Rodrigues-Luiz, Bruna M. Valente, Antonio Edson R. Oliveira, Rafael B. Polidoro, Luciana W. Zuccherato, Daniella C. Bartholomeu, Santuza M. R. Teixeira.

Software: A. Trey Belew, Gabriela F. Rodrigues-Luiz, Antonio Edson R. Oliveira, Daniella C. Bartholomeu.

Supervision: Santuza M. R. Teixeira.

Validation: A. Trey Belew, Caroline Junqueira, Bruna M. Valente, Daniella C. Bartholomeu, Najib M. El-Sayed, Santuza M. R. Teixeira.

Visualization: A. Trey Belew.

Writing – original draft: Najib M. El-Sayed, Santuza M. R. Teixeira.

Writing – review & editing: Caroline Junqueira, Sergio Schenkman, Ricardo T. Gazzinelli, Barbara A. Burleigh.

References

1. Brener Z. Biology of Trypanosoma cruzi. Annu Rev Microbiol. 1973; 27:347–82. https://doi.org/10.1146/annurev.mi.27.100173.002023 PMID: 4201691
2. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran A-N, et al. The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science. 2005; 309(5733):409–15. https://doi.org/10.1126/science.1112631 PMID: 16020725
3. Weatherly DB, Boehlke C, Tarleton RL. Chromosome level assembly of the hybrid Trypanosoma cruzi genome. BMC Genomics. 2009; 10:255. https://doi.org/10.1186/1471-2164-10-255 PMID: 19486522
4. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. 2010; 38:D457–62. https://doi.org/10.1093/nar/gkp851 PMID: 19843604
5. Buscaglia CA, Di Noia JM. Trypanosoma cruzi clonal diversity and the epidemiology of Chagas' disease. Microbes Infect. 2003; 5(5):419–27. PMID: 12737998
6. Zingales B, Miles M a, Campbell D a, Tibayrenc M, Macedo AM, Teixeira MMG, et al. The revised Trypanosoma cruzi subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect Genet Evol. 2012; 12(2):240–53. https://doi.org/10.1016/j.meegid.2011.12.009 PMID: 22226704
7. Reis-Cunha JL, Rodrigues-Luiz GF, Valdivia HO, Baptista RP, Mendes TAO, de Morais GL, et al. Chromosomal copy number variation reveals differential levels of genomic plasticity in distinct Trypanosoma cruzi strains. BMC Genomics. 2015; 16(1).
8. Franzen O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA, et al. Shotgun sequencing analysis of Trypanosoma cruzi I Sylvio X10/1 and comparison with T. cruzi VI CL Brener. PLoS Negl Trop Dis. 2011; 5(3):e984. https://doi.org/10.1371/journal.pntd.0000984 PMID: 21408126
9. Araújo PR, Teixeira SM. Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in Trypanosoma cruzi: a review. Mem Inst Oswaldo Cruz. 2011; 106(3):257–66. PMID: 21655811
10. Minning T a, Weatherly DB, Atwood J, Orlando R, Tarleton RL. The steady-state transcriptome of the four major life-cycle stages of Trypanosoma cruzi. BMC Genomics. 2009; 10:370. https://doi.org/10.1186/1471-2164-10-370 PMID: 19664227

11. Li Y, Shah-Simpson S, Okrah K, Belew AT, Choi J, Caradonna KL, et al. Transcriptome Remodeling in Trypanosoma cruzi and Human Cells during Intracellular Infection. PLoS Pathog. 2016; 12(4): e1005511. https://doi.org/10.1371/journal.ppat.1005511 PMID: 27046031
12. Berna L, Chiribao ML, Greif G, Rodriguez M, Alvarez-Valin F, Robello C. Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in Trypanosoma cruzi. PeerJ. 2017; 5:e3017. https://doi.org/10.7717/peerj.3017 PMID: 28286708
13. Houston-Ludlam GA, Belew AT, El-Sayed NM. Comparative Transcriptome Profiling of Human Foreskin Fibroblasts Infected with the Sylvio and Y Strains of Trypanosoma cruzi. PLoS One. 2016; 11(8): e0159197. https://doi.org/10.1371/journal.pone.0159197 PMID: 27505626
14. Lima MT, Jansen AM, Rondinelli E, Gattass CR. Trypanosoma cruzi: properties of a clone isolated from CL strain. Parasitol Res. 1991; 77(1):77–81. PMID: 1899726
15. Sousa MA, Pereira SM. Cryptic infections in mice with the trypanosoma cruzi CL-14 clone. Rev Inst Med Trop Sao Paulo. 1999; 41(3):205–6. PMID: 10529843
16. Lima MT, Lenzi HL, Gattass CR. Negative tissue parasitism in mice injected with a noninfective clone of Trypanosoma cruzi. Parasitol Res. 1995; 81(1):6–12. PMID: 7724515
17. Junqueira C, Santos LI, Galvao-Filho B, Teixeira SM, Rodrigues FG, DaRocha WD, et al. Trypanosoma cruzi as an effective cancer antigen delivery vector. Proc Natl Acad Sci USA. 2011; 108(49):19695–700. https://doi.org/10.1073/pnas.1110030108 PMID: 22114198
18. Paiva CN, Pyrrho AS, Ribeiro LJ, Goncalves R, Costa DA, Araujo-Jorge TC, et al. Trypanosoma cruzi: requirements for induction and maintenance of protective immunity conferred by immunization. Exp Parasitol. 2002; 102(2):89–98. PMID: 12706744
19. Soares MB, Goncalves R, Pyrrho AS, Costa DA, Paiva CN, Gattass CR. Balanced cytokine-producing pattern in mice immunized with an avirulent Trypanosoma cruzi. An Acad Bras Cienc. 2003; 75(2):167–72. PMID: 12894302
20. Maugeri DA, Cannata JJ, Cazzulo JJ. Glucose metabolism in Trypanosoma cruzi. Essays Biochem. 2011; 51:15–30. https://doi.org/10.1042/bse0510015 PMID: 22023439
21. Mucci J, Lantos AB, Buscaglia CA, Leguizamon MS, Campetella O. The Trypanosoma cruzi Surface, a Nanoscale Patchwork Quilt. Trends Parasitol. 2017; 33(2):102–12. https://doi.org/10.1016/j.pt.2016.10.004 PMID: 27843019
22. Queiroz RML, Charneau S, Mandacaru SC, Schwämmle V, Lima BD, Roepstorff P, et al. Quantitative Proteomic and Phosphoproteomic Analysis of Trypanosoma cruzi Amastigogenesis. Mol Cell Proteomics. 2014; 13(12):3457–72. https://doi.org/10.1074/mcp.M114.040329 PMID: 25225356
23. De Gaudenzi J, Frasch AC, Clayton C. RNA-binding domain proteins in Kinetoplastids: a comparative analysis. Eukaryot Cell. 2005; 4(12):2106–14. https://doi.org/10.1128/EC.4.12.2106-2114.2005 PMID: 16339728
24. Guerra-Slompo EP, Probst CM, Pavoni DP, Goldenberg S, Krieger MA, Dallagiovanna B. Molecular characterization of the Trypanosoma cruzi specific RNA binding protein TcRBP40 and its associated mRNAs. Biochem Biophys Res Commun. 2012; 420(2):302–7. https://doi.org/10.1016/j.bbrc.2012.02.154 PMID: 22425988
25. Caro F, Bercovich N, Atorrasagasti C, Levin MJ, Vazquez MP. Trypanosoma cruzi: analysis of the complete PUF RNA-binding protein family. Exp Parasitol. 2006; 113(2):112–24. https://doi.org/10.1016/j.exppara.2005.12.015 PMID: 16460732
26. Kramer S, Kimblin NC, Carrington M. Genome-wide in silico screen for CCCH-type zinc finger proteins of Trypanosoma brucei, Trypanosoma cruzi and Leishmania major. BMC Genomics. 2010; 11:283. https://doi.org/10.1186/1471-2164-11-283 PMID: 20444260
27. Caro F, Bercovich N, Atorrasagasti C, Levin MJ, Vazquez MP. Protein interactions within the TcZFP zinc finger family members of Trypanosoma cruzi: implications for their functions. Biochem Biophys Res Commun. 2005; 333(3):1017–25. https://doi.org/10.1016/j.bbrc.2005.06.007 PMID: 15964555
28. Hendriks EF, Matthews KR. Disruption of the developmental programme of Trypanosoma brucei by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein. Mol Microbiol. 2005; 57(3):706–16. https://doi.org/10.1111/j.1365-2958.2005.04679.x PMID: 16045615
29. Lott K, Mukhopadhyay S, Li J, Wang J, Yao J, Sun Y, et al. Arginine methylation of DRBD18 differentially impacts its opposing effects on the trypanosome transcriptome. Nucleic Acids Res. 2015; 43 (11):5501–23. https://doi.org/10.1093/nar/gkv428 PMID: 25940618
30. Erben ED, Fadda A, Lueong S, Hoheisel JD, Clayton C. A genome-wide tethering screen reveals novel potential post-transcriptional regulators in Trypanosoma brucei. PLoS Pathog. 2014; 10(6):e1004178. https://doi.org/10.1371/journal.ppat.1004178 PMID: 24945722
31. Archer SK, Luu VD, de Queiroz RA, Brems S,Niyogi S, Mucci J, Campetella O, Docampo R. Rab11 Regulates Trafficking of Trans-sialidase to the Plasma Membrane through the Contractile Vacuole

Complex of Trypanosoma cruzi. PLoS Pathog. 2014 Clayton C. Trypanosoma brucei PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. PLoS Pathog. 2009; 5(8): e1000565.

32. Gunzl A. The pre-mRNA splicing machinery of trypanosomes: complex or simplified? Eukaryot Cell. 2010; 9(8):1159–70. https://doi.org/10.1128/EC.00113-10 PMID: 20581293

33. Freitas LM, dos Santos SL, Rodrigues-Luiz GF, Mendes TAO, Rodrigues TS, Gazzinelli RT, et al. Genomic analyses, gene expression and antigenic profile of the trans-sialidase superfamily of trypanosoma cruzi reveal an undetected level of complexity. PLoS One. 2011; 6(10).

34. DaRocha WD, Silva RA, Bartholomeu DC, Pires SF, Freitas JM, Macedo AM, et al. Expression of exogenous genes in Trypanosoma cruzi: improving vectors and electroporation protocols. Parasitol Res; 2004; 92(2):113–20. https://doi.org/10.1007/s00436-003-1004-5 PMID: 14634799

35. Rubin-de-Celis SSC, Uemura H, Yoshida N, Schenkman S. Expression of trypomastigote trans-sialidase in metacyclic forms of Trypanosoma cruzi increases parasite escape from its parasitophorous vacuole. Cell Microbiol. 2006; 8(12):1888–98 https://doi.org/10.1111/j.1462-5822.2006.00755.x PMID: 16824037

36. Niyogi S, Mucci J, Campetella O, Docampo R. Rab11 Regulates Trafficking of Trans-sialidase to the Plasma Membrane through the Contractile Vacuole Complex of Trypanosoma cruzi. PLoS Pathog. 2014; 10(6):e1004224 https://doi.org/10.1371/journal.ppat.1004224 PMID: 24968013

37. Atayde VD, Neira I, Cortez M, Ferreira D, Freymuller E, Yoshida N. Molecular basis of non-virulence of Trypanosoma cruzi clone CL-14. Int J Parasitol. 2004; 34(7):851–60. https://doi.org/10.1016/j.ijpara.2004.03.003 PMID: 15157768

38. Tonelli RR, Silber AM, Almeida-de-Faria M, Hirata IY, Colli W, Alves MJM. l-Proline is essential for the intracellular differentiation of Trypanosoma cruzi. Cell Microbiol. 2004; 6(8):733–41. https://doi.org/10.1111/j.1462-5822.2004.00397.x PMID: 15236640

39. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

40. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25(9):1105–11. https://doi.org/10.1093/bioinformatics/btp120 PMID: 19289445

41. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

42. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2014; 31(2):166–9. https://doi.org/10.1093/bioinformatics/btu638 PMID: 25260700

43. Leek JT, Storey JD. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. PLoS Genet. 2007; 3(9):e161.

44. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. Biostatistics. 2011; 13(3):539–52. https://doi.org/10.1093/biostatistics/kxr034 PMID: 22101192

45. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43(7):e47–e47. https://doi.org/10.1093/nar/gkv007 PMID: 25605792

46. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15(12).

47. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009; 26(1):139–40. https://doi.org/10.1093/bioinformatics/btp616 PMID: 19910308

48. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010; 11(2):R14. https://doi.org/10.1186/gb-2010-11-2-r14 PMID: 20132535

49. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, et al. g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 2016; 44(W1):W83–9. https://doi.org/10.1093/nar/gkw199 PMID: 27098042

Apêndice IV: Manuscrito a ser submetido na revista Journal of Biological Chemistry contendo dados de parasitas nocautes para a TcRBP99.

A novel RNA Binding Protein that controls gene expression in epimastigotes and affects metacyclogenesis of *Trypanosoma cruzi*

Bruna Mattioly Valente[1], Wanessa Moreira Goes[1], Thais Silva Tavares[1], Viviane G. Silva[1], Trey A. Bellew[3], Fabiano Sviatopolk-Mirsky Pais[2], Najib M. El-Sayed[3] and Santuza M.R. Teixeira[1]*

[1]Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil, [2]Centro de Pesquisas Rene Rachou, Fundação Oswaldo Cruz, Belo Horizonte, MG, Brazil, [3]Department of Cell Biology and Molecular Genetics and Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD, USA,

* Corresponding author

santuzat@ufmg.br

ABSTRACT

Trypanosoma cruzi, the causative agent of Chagas disease, has three biochemically and morphologically distinct developmental stages that are programed to rapidly respond to drastic environmental changes the parasite faces during its life cycle. Unlike other eukaryotes, the *T. cruzi* genome contains protein coding genes that are transcribed into polycistronic pre-mRNAs before they are processed into mature mRNAs through coupled “trans-splicing” and poly-adenylation reactions. Because of this, control of gene expression relies mainly on post-transcriptional mechanisms that mainly affect steady-state levels and translation rates of mRNAs. Transcriptome analyses comparing *in vitro* cultured epimastigotes, trypomastigotes and intracellular amastigotes revealed changes in gene expression that reflect the parasite adaptation to distinct environments, including changes in energy sources, oxidative stress responses, cell cycle control and cell surface components. These analyses also revealed significant changes in the expression of genes encoding RNA binding proteins (RBP), considered the main trans-acting factors responsible for post-transcriptional control of gene expression. Here, we characterized a gene encoding a RBP containing a zinc finger motif, named TcRBP99, whose transcript levels are up-regulated 25-fold in epimastigotes compared to trypomastigotes and amastigotes. Its role in controlling gene expression in the insect stage of the parasite life cycle was revealed by RNA-seq analysis of epimastigote cell lines in which the TcRBP99 gene was disrupted: transcript levels of 11 genes, all of them being up-regulated in wild type epimastigotes, were significantly reduced in knockout epimastigotes compared to wild type parasites. Among these targets of TcRBP99 are genes encoding amino acid transporters and a gene encoding a

protein associated with parasite differentiation. Since knockout mutant presented reduced proliferation rates and increased capacity to differentiate into metacyclic trypomastigotes compared to wild type parasites, we suggest that this RBP not only positively regulates the expression of genes involved with epimastigote survival and proliferation in the insect vector, but also acts as a negative regulator of metacyclogenesis.

Key words: gene expression, RNA binding protein, Trypanosoma cruzi, post-transcriptional regulation

INTRODUCTION

Trypanosoma cruzi causes Chagas disease, a neglected parasitic infection found in Latin America, where over 8 million people are affected. About 30% of infected individuals develop serious clinical symptoms such as cardiomyopathy or severe digestive tract disorders, resulting in approximately 14,000 deaths annually. No vaccines are available to prevent infection and the two current drugs used to treat infected people, besides being effective only during the chronic phase of Chagas disease, have serious adverse side effects (http://www.who.int/topics/chagas_ disease/en). *T. cruzi* transmission to humans occurs mostly by domiciliated triatomine vectors, being *Triatoma infestans* and *Rhodnius prolixus* the main Chagas disease vectors. Although recent vector control programs have succeeded in reducing infection rates, non-vectorial transmission routes, including transmission via contaminated fruit products, blood transfusions and congenital transmission still frequently occur. When taking a blood meal on an infected mammal, the insect vector takes up trypomastigotes circulating in the blood, which, once inside the insect hind gut, differentiate into replicative epimastigotes. In the posterior end of the digestive tract, epimastigotes differentiate into infective, non-dividing metacyclic trypomastigotes, which can be expelled with the vector's feces during a blood meal. If inoculated through ruptures in the skin of a new mammalian host, trypomastigotes can circulate in the bloodstream and infect different cell types, including macrophages. Once in the host cell cytoplasm, trypomastigotes differentiate into replicative amastigotes, which, after replicating for 3-5 days, differentiate again into highly motile trypomastigotes that lyse the host cell membrane. Newly release trypomastigotes

reach the circulatory system and propagate the infection by invading other cells or when they are taken up by the insect vector [1].

The existence of several biochemically and morphologically distinct stages that alternate between an invertebrate and a vertebrate host implies that this parasite must have a genetic developmental program that can rapidly respond to the drastic environmental changes it faces during its life cycle. Like other members of the Trypanosomatid family, genome studies revealed that all *T. cruzi* protein coding genes are organized into long polycistronic transcription units that are transcribed into polycistronic pre-mRNAs, which are subsequently processed into mature, monocistronic mRNAs through coupled “trans-splicing” and poly-adenylation reactions [2]. Because of polycistronic transcription, control of gene expression relies mainly on post-transcriptional mechanisms that mainly affect steady-state levels and translation rates of mRNAs [3]. Recent transcriptome studies performed with distinct parasite strains [4-7] have showed that in average 30-50% of the predicted protein coding genes undergo significant changes in gene expression throughout the parasite life cycle. These studies also showed that large gene families involved with remodeling the parasite cell surface account for the most substantial changes observed not only when distinct parasite life stages were compared [4-6] but also when parasite strains showing distinct virulence phenotypes were analyzed [7].

Consist with their role as key elements involved with post-transcriptional control of gene expression, a large variety of RNA Binding Proteins (RBP) are encoded in the *T. cruzi* genome [8-11]. Through direct or indirect RNA-protein interactions, RBPs are responsible for the control of steady-state levels and translation rates of their target mRNAs in many different cell types [12, 13]. Recent analyses based

116 on RNA-seq data revealed that among the 147 *T. cruzi* genes present in the
117 parasite genome, only a small proportion of them showed significant variations in
118 their expression when distinct mammalian stages of the parasite life cycle are
119 compared. The Y strain transcriptome data revealed eight genes encoding RBPs
120 whose expression levels significantly differ during the intracellular infection
121 process (logFC>1 and P-value < 0.05) [4]. Transcriptome analyses performed
122 with the CL Brener strain identified six differentially expressed genes (logFC>2
123 and P-value < 0.05) when amastigotes harvest at 60 and 96 hours pos-infection
124 and extracellular trypomastigotes were compared. However, none of these
125 studies have compared the expression levels of genes encoding RBPs in all
126 parasite life cycle stages including epimastigotes. Among the reports describing
127 *T. cruzi* RBPs, there are reports about a zinc finger RBP involved with
128 metacyclogenesis [14], a Pumilio-like protein that is part of the mRNA
129 degradation pathway [15] and an RBP that belongs to the RNA-recognition motif
130 family (RRM), named TcRBP19, that is detected only in the intracellular
131 amastigote stage [16]. TcRBP40 is an RBP constitutively expressed throughout
132 the *T. cruzi* life cycle, which shows distinct cellular localization patterns when the
133 three forms of the parasite were compared [9]. TcAlba30 is also constitutively
134 expressed but is notable for interacting with β-amastin mRNAs, which are highly
135 abundant transcripts present in epimastigotes [17]. However, despite the large
136 number of examples pointing towards the importance of RBPs in the post-
137 transcriptional control of mRNA levels and translational efficiency in *T. cruzi* as
138 well as several other trypanosomatid species, only one report provided direct
139 experimental evidence, derived from knockout cell lines, of the role of a RBP in
140 controlling gene expression in *T. cruzi*. Alcantara and cols. have recently showed

that knockout epimastigote cell lines of a RBP that is exclusively expressed in the insect stages of the parasite are incapable to differentiate into the metacyclic trypomastigotes [18]. One of the main goals of the work presented here was to perform a global gene expression analyses comparing epimastigotes of the CL Brener cloned strain with the mammalian stages of the *T. cruzi* to identify RBPs with the highest expression levels in epimastigotes. Once identified, we proceed to the characterization of one of these proteins based on the assumption that an RBP that is highly expressed in epimastigotes may act as a key regulatory element controlling gene expression during the insect stage of *T. cruzi*. The analyses of parasite cell lines over-expressing the gene encoding this RBP as well as analyses of knockout parasite cell lines allowed us to describe a novel *T. cruzi* RBP which is involved with the expression of epimastigote specific genes and acts as a negative regulator of metacyclogenesis.

METHODS

Parasites cultures

Epimastigotes forms of *T. cruzi* CL Brener clone were maintained at 28ºC in liver infusion tryptose medium (LIT) supplemented with 10% of fetal bovine serum as previous described [19].

In vitro T. cruzi metacyclogenesis

The metacyclogenesis was performed under stress conditions as described by [20]. For this, CL Brener epimastigotes forms were maintained in LIT medium in

exponential-phase and total of 2x'07 epimastigotes was transferred to RPMI medium without fetal bovine serum. The cells were left undisturbed for four days and Proportion of metacyclic trypomastigotes were evaluated every two days.

In silico analyses of T. cruzi RNA binding proteins

All *T. cruzi* transcripts containing RNA binding motives, previously identify by [7], were analyzed throughout the life cycle of this parasite using RNA-seq data to identify differentially expressed genes in epimastigotes forms.

Proteins sequences orthologous of RBP with differential expression in epimastigotes forms were extract from different organisms in TritypDB database. The multiple alignments sequences were performed using Muscle algorithm (http://www.ebi.ac.uk/Tools/msa/muscle/) and phylogenetic three was built using MEGA software.

Over expression and cellular localization of an HA-tagged T. cruzi RBP

Complete coding sequence of TcCLB.506739.99 were amplified by PCR from genomic CL Brener strain DNA, using forward For_TcRBP99_XbaI and reverse Rec_TcRBP99::HA_XhoI primers containing *Xba*I and *Xho*I restriction sites respectively (TableS1). Sequence corresponding to HA epitope was introduced in C-terminal of the TcRBP39.99 sequence. Amplicons were purified and inserted in pROCKGFPNeo [21] expressing vector, previously digested with *Xba*I and *Xho*I, generating pROCK.RBP39.99:HA plasmid. For transfection 20 µg of this plasmid, digested with *Not*I restriction enzyme, and $4x10^7$ epimastigotes forms

were resuspended in 100 µl of nucleofector buffer [22] and submitted to the program U033 in AMAXA Nucleofactor. Twenty-four hours after transfection, parasites were selected with G418 (200 µg/ml) antibiotic. Clones were obtained by limiting dilution platting 0.5 parasites/well. Parasites expressing TcRBP39.99 with HA tag were confirmed by Western Blot assay, where $1x10^6$ cells were load in SDS-polyacrylamide gel followed by incubation of anti-HA (Sigma) and anti-IgG (Sigma) diluted 1:5000 and 1:3000 respectively. For cellular localization, epimastigotes forms expressing the TcRBP39.99 with HA tag were fixed with 4% paraformaldehyde for 20 min at 4ºC, blocked and permeabilized with block solution (1% BSA and 0.2% Triton X-100) for 20 min at room temperature. Then, the parasites were incubated with 1:50 anti-HA (Roche) for 45 min and secondary anti-Rat IgG, diluted 1:250, conjugated to Alexa Fluor 488 (Molecular Probes/ Life Technologies) for 30 min. After tree washes with PBS1, 1 µg/ml of DAPI (Molecular Probes/ Life Technologies) were used to stain the nuclei and cover slides were mounted with prolong gold anti-fade (Molecular Probes/ Life Technologies). Imagens were captured on Optical fluorescence microscopy Axio Imager Z2 – ApoTome 2-Zeiss.

Generation of TcRBP39.99 knockout cell lines

DNA constructs were generated to delete both TcRBP39.99 alleles by homologous recombination in epimastigotes forms of *T. cruzi* CL Brener clone. For this, PCR amplifications corresponding to 5' (from 150pb before the first ATG until 250pb after the ORF begin) and 3' (from 200pb before the stop codon until 200pb after the stop codon) sequences of this gene were realized (for primers sequences see TableS1). Restriction sites for *Hind*III/*Sac*I and *Xho*I/*Xba*I

212 enzymes were added in 5' and 3' respectively. PCR product corresponding to the
213 5' was first cloned upstream of pTopo_HX1_Neo_GAPDH [23] and then the 3'
214 fragment was cloned downstream of GAPHD region, generating the plasmid
215 TcRBP39.99_Neo. To generate parasites single knockout, this plasmid were
216 used as templet to PCR with the primers For5'KO_TcRBP99_HindIII and
217 Rev3'KO_TcRBP99_XbaI and PCR products were used to transfect parasites
218 wild type. Twenty-four hours post transfection the parasites single knockouts
219 were selected with G418 (200 µg/ml) antibiotic and clones were obtained by
220 limiting dilution platting 0.5 parasites/well. The deletion of the second allele the
221 plasmid TcRBP39.99_Neo were digest with *Spe*I and *Not*I restriction enzymes
222 and insert to plasmid pTopo_HX1_Hygro_GAPDH previously digest with the
223 same enzymes, generating the plasmid TcRBP99_Hygro. PCR products
224 generate by amplification with the primers For5'KO_TcRBP99_HindIII and
225 Rev3'KO_TcRBP99_XbaI were used to transfect one single knockdown clone to
226 the TcRBP39.99. After twenty-four hours hygromicin (200 µg/ml) was added to
227 selected population of knockdown parasites and clones were obtained by limiting
228 dilution platting 0.5 parasites/well.

229

230 RNA extraction and quantitative PCR analyses

231 Total RNA was extraction from 10^8 epimastigotes from exponentially growing
232 cultures using Illustra RNAspin Mini kit (GE Healthcare). cDNA synthesis were
233 generated using SuperScript II Reverse Transcriptase kit first-strand synthesis
234 (Invitrogen) and Oligo$(dT)_{18}$, according to the manufacturer's instructions.
235 Specific primers to TcRBP99 gene (TableS1) and SYBR Green (Bio-Rad) were

used to quantify PCR products, in the reaction containing 0.5 µM of each primer, 1X SYBR Green and 1µl of cDNA diluted 1:32 in total of 10µl. Applied Biosystems 7900HT Fast Real-Time PCR System (Life Technologies) platform was used to run the reactions. Relative levels of mRNAs were normalized with specifics primers to amplify the gene encoding for constitutive 60S ribosomal protein L9 (TcCLB.504181.10), see Table1 for primers sequences, and quantify using standard curve method based on cycle threshold values ($2^{-\Delta\Delta ct}$) [24].

RNA sequencing and bioinformatics analysis

The quality of RNA were determined by Agilent 2100 bioanalyzer and quantified by qPCR using a KAPA Biosystems library quantification kit. RNA coming from duplicates of two clones TcRBP39.99 null-mutants and triplicates from wild type *T. cruzi* CL Brener clone were submitted to Illumina® TruSeq® RNA Sample Preparation Kit v2 to generate RNA-Seq libraries. Illumina next-generation sequencing were performed using MiSeq platform (at Instituto René Rachou – Fiocruz Minas). Raw reads quality were evaluated using the tool FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimming with Trimmomatic software [25]. Sequences were mapped in *T. cruzi* CL Brener genome version 9.0 (available on www.tritrypdb.org) using TopHat2 software with the parameters to allow 2 mismatches and the number maximum of multihits was setting for 20. To count the number of reads align in each coding sequencing (CDS) was employed the software HTSeq [26] and DESeq2 package was used to analyze genes with differential expression, defined as genes with log2>2 or log2<2 and q-value<0.05, between wild type and null mutants.

RNA immunoprecipitation

Epimastigotes expressing TcRBP99 with HA tag were lysed in lysis buffer (Tris-HCl pH7.5 10mM; NaCl 140mM; NP-40 1%), centrifugated and the supernatant were incubated with bead EZview Red Anti-HA (Sigma – Aldrich). After 16 hours the beads were centrifugated and the flow through was transferred to new tube. The beads were washed with buffer lysis and eluted in elution buffer (62,5 mM Tris-HCl pH 6,8; 10% glycerol; 2% SDS; 5% β-mercaptoethanol; 0,002% bromophenol blue).

RESULTS AND DISCUSSION

Differential gene expression analyses identified a novel *T. cruzi* RNA binding protein up-regulated in epimastigotes.

Previous RNA-seq analyses comparing trypomastigotes and intracellular amastigotes from the CL Brener cloned strain of *T. cruzi* harvested at different time points post-infection of human fibroblasts revealed large differences in gene expression that reflect the parasite adaptation to the mammalian host. Here we compared RNA-seq data from mRNA obtained from *in vitro* cultivated epimastigotes with the RNA-seq data derived from tissue culture derived trypomastigotes and intracellular amastigotes. DE-Seq analyses showed that, whereas in trypomastigotes highly expressed genes encode cell surface proteins and in amastigotes, highly expressed genes are involved with transport of nutrients from the host cell cytoplasm, epimastigotes show high levels of transcripts from genes involved with amino acid transporters and metabolism, proteolysis as well as mRNAs encoding for beta amastins and a protein

associated with differentiation. Gene ontology (GO) enrichment analysis performed from pairwise comparisons between epimastigotes, trypomastigote and amastigotes collected at 60 hpi, showed that genes involved with oxidative-reduction process, amino acid metabolism and respiratory chain are among the highly up-regulated genes in epimastigotes (Table 1). These results are in agreement with the findings recently described by Berna et.al, [5] who showed that, besides genes involved with protein degradation and amino acid metabolism, genes encoding enzymes of the Krebs cycle and oxidative phosphorylation as well as steroid biosynthesis are among the genes that are up-regulated in epimastigotes of the Dm28c strain [5].

Because post-transcriptional control mechanisms mediated by RNA binding proteins (RBP) are largely responsible for controlling gene expression in *T. cruzi*, we determined which RBP genes presented the largest difference in transcript levels when epimastigotes were compared to the other two forms of the parasite life cycle. As described previously, a total of 253 *T. cruzi* sequences representing 147 genes encoding RBPs (since two alleles were identified for most genes) were previously identified in the *T. cruzi* CL Brener genome [7]. Among those, RNA-seq analyses showed that transcripts from five genes were significantly up-regulated in epimastigotes comparing to amastigotes collected at 60 hours post-infection and trypomastigotes collected from the supernatant of infected cells (Fig 1A). The genes encoding differentially expressed RBPs identified in these analyses were TcCLB.506739.99, TcCLB.506931.4 and TcCLB.510134.44, which contains a zinc finger domain and TcCLB.511727.270 and TcCLB.511727.290, which have two RRM motives in their sequences. Umaer and Williams (2015) have characterized these two RBPs containing the RRM

motifs as proteins involved with ribosome biogenesis that are exclusive from trypanosomatids. The TcCLB.506931.4 and TcCLB.510131.44 genes have been characterized as hypothetical protein. Among the differentially expressed genes encoding RBPs, the TcCLB.506739.99 gene showed the largest difference in gene expression when epimastigotes were compared to tissue culture derived trypomastigotes (8.9-fold) and amastigotes (7.6-fold) (Fig 1A-B). In the genome database (http://tritrypdb.org/tritrypdb/), TcCLB.506739.99 is annotated as a gene encoding a 19.5 KDa zinc finger containing RBP that is highly expressed in epimastigotes as indicated by microarray analyses, performed with the Brazil strain, but no further characterization of this gene was described. Because of its increased expression in epimastigotes, we hypothesized that this RBP, named TcRBP99, may be involved with the control of epimastigote-specific gene expression.

Due to the hybrid nature of the CL Brener genome, two alleles for this gene were identified with the correspondent non-esmeraldo allele, TcCLB.510819.119, encoding a similar, 173 amino acid protein with 97.7% amino acid identity. According to microarray data found in tritrypdb, transcripts derived from the second allele also presented increased levels in epimastigotes compared to trypomastigotes and amastigotes (3.5-fold and 2.1-fold, respectively). According to our RNA-seq analyses, the differences of transcripts that mapped to the TcCLB.510819.119 allele are 2.16 and 2.04-fold when epimastigotes are compared to 60 hpi amastigotes and trypomastigotes, respectively. Similarly, increased transcript levels in epimastigotes were also determined using RNA-seq data derived from Y and Dm28c strains provided by Li and cols (2016) and Berna and cols (2015) [4, 5].

Searches in the database revealed several orthologues of the TcRBP99 gene in different trypanosomatid species, but no similar sequences were found in the genomes of other eukaryotes. Multiple sequence alignment revealed a high degree of conservation at the N-terminal region (Fig 2A), where is localized the zinc finger domain (C-X8-C-X5-C-X3-H). A phylogenetic tree constructed with sequences derived from various members of the Trypanosomatid family showed that amino acid sequences fromTcRBP99 homologues can be clustered in 3 sub-families: a sub-family comprising *T. cruzi* and *T. rangeli* sequences, a sub-family with *T. brucei*, *T. evansi*, *T. vivax* and *T. congolensis* sequences and a third sub-family that includes sequences derived from Leishmania, Leptomonas and Crithidia. (Fig 2B). By transfecting epimastigotes with an HA-tagged sequence of TcRBP99 (see next session), we were able to determine that this protein has a cytoplasmic localization (Fig 2C). Thus, besides the presence of the zinc finger domain, the high isoelectric point of its polypeptide chain (8.49) and cytoplasmic localization all point towards a role of TcRBP99 as an RNA binding factor.

Overexpression of TcRBP99 resulted in an inhibition of *T. cruzi* metacyclogenesis whereas TcRBP99 knockouts showed increased metacyclogenesis capacity

To investigate the role of TcRBP99 as a factor that may be involved in post-transcriptional regulation during the *T. cruzi* life cycle, we analyzed parasite cell lines over-expressing this protein. A 20 KDa protein was expressed in epimastigotes transfected with the pROCK-Neo vector containing the HA-tagged coding region of the TcCLB.506739.99 gene under the control of the ribosomal promoter and 5’ and 3’ UTR sequences from constitutively expressed genes (Fig

3A-B). Quantitative qPCR analyses showed that transfected parasites have a 1.5-fold increased expression of TcRBP99 transcripts compared to non-transfected parasites (Fig 3C). Growth curves showed that the over-expression of TcRBP99 increased growth rates of epimastigotes, since after five days in culture, a higher cell density was observed with the transfected parasites compared with wild type cells and parasites transfected with empty pROCK-Neo vector (Fig 3D). When epimastigotes were transferred to a medium (RPMI) that induces epimastigotes differentiation into metacyclic trypomastigotes (metacyclogenesis) a clearly distinct phenotype was observed with parasites over-expressing TcRBP99: whereas in cultures of wild type epimastigotes or epimastigotes transfected with the empty vector about 15-20% of the cells differentiated into metacyclic trypomastigotes 10 days after being transferred to the RPMI medium, only 10% of the cells over-expressing TCRBP99 showed the metacyclic morphology after the same time under differentiation conditions (Fig 3E). This result suggests that the TcRBP99 is a protein that positively regulates proliferation of epimastigotes and negatively regulates parasite genes required for the differentiation from replicative epimastigotes into non-replicative and infective metacyclic trypomastigotes.

To further investigate the role of TcRBP99 as a regulatory RNA binding protein we generated knockout cell lines in which one allele was disrupted with the Neomycin gene and the second allele with the Hygromycin gene (Fig 4A). We confirmed the correct insertion of the Neomycin resistance cassette into the first allele, generating a heterozygous mutant, as shown by PCR products using primers P1 and P2, whereas PCR products using primers P3 and P4 confirmed the insertion of the Hygromycin resistance cassette into the second allele. PCR

using primers P5 and P6 also confirmed the disruption of both alleles by the drug resistance cassettes (Fig 4B). As shown in the right panels of Fig 4A-B, using the pROCK vector containing the puromycin resistance gene as a third drug selection marker, we generated an addback epimastigote cell line, in which the sequence corresponding to TcCLB.506739.99 gene was re-inserted at the tubulin locus in the parasite genome of one cloned cell line derived from the knockout parasite population. Expression of the TcRBP99 gene was evaluated in all mutant cell lines using quantitative qPCR: as shown in Fig 4C, heterozygous mutants have 50% less transcripts encoding this gene whereas no transcripts were detected in both null mutant cell lines. Because the addback cell line was generated using the pROCK vector, which has the strong ribosomal promoter controlling the expression of the transfected gene, a 2.5-fold increase in transcript levels of the gene was detected in this cell line comparing with wild type parasites.

Growth curve analyses of two cloned cell line derived from the knockout mutant population showed a growth defect in both knockout clones compared to wild type epimastigotes. The differences in the numbers of cells, which are more pronounced at days 7 and 9 of culture, are reduced when we compared wild type cultures with the two cloned cell lines derived from the addback population (Fig 4D). The results from growth curve analyses indicate that epimastigote proliferation is affected by the knockout of the TcCLB.506739.99 gene. More importantly, by determining the rates of metacyclogenesis, we discovered that the lack of TcRBP99 resulted in increased differentiation from epimastigotes to metacyclic trypomastigotes (Fig 4E): both cloned knockout cell lines showed a minimum of 2-fold increase in metacyclogenesis rates compared to wild type cells, a difference that can be observed as early as 4 days after transferring

parasites to the RPMI medium. Adding back the TcRBP99 gene not only restored the normal growth rate but also the normal differentiation capacity of epimastigotes: on day 8 after being transfer to the RPMI medium, both addback cell lines showed similar rates of metacyclogenesis compared to wild type cells (Fig 4E). Taken together, the distinct phenotypes observed with TcRBP99 knockout and over-expressing parasites corroborate with the assumption that TcRBP99 is a regulatory factor that controls the expression of epimastigote genes involved with epimastigote proliferation and negatively regulates the expression of genes required for the induction of metacyclogenesis.

TcRBP99 knockout resulted in inhibited expression of *T. cruzi* genes that are up-regulated in epimastigotes

As shown by Berna et.al, (2017) and the RNA-seq analyses described here, the top most up-regulated transcripts expressed in epimastigotes encode proteins involved with amino acid metabolism and oxidative phosphorylation. Using RNA-seq analyses, we asked which genes have their expression affected by TcRBP99 knockout cell lines. Total RNA was isolated from wild type epimastigotes as well as from two cloned knockout mutant cell lines and mRNA populations derived from replicates of each sample were sequenced on the Illumina platform. Table 2 showed the total number of reads generated from all libraries as well as the percentages of reads that mapped to the *T. cruzi* CL Brener reference genome. Mapped sequencing data derived from all libraries were analyzed using principal component analysis (PCA) to inspect the relationships between samples and to identify potential outliers. Supplementary Fig S1 showed that the global transcriptome profiles of wild type epimastigotes significantly differ from the

profiles of both knockout mutants, indicating that the disruption of the TcRBP99 gene affects gene expression in epimastigotes. A list of 11 genes whose expression levels were significantly affected (p-value < 0.05 and log2FC >1) by TcRBP99 knockout is shown in Table 3. It is interesting to note that only genes whose expression was down-regulated in the knockout cell lines compared to wild type epimastigotes were identified. Since no significant changes resulting in increased expression of any gene in the knockout parasites were found, we concluded that TcRBP99 acts mainly as a positive regulatory factor controlling gene expression. It is also important to note that the RNA-seq analyses comparing epimastigotes, trypomastigotes and intracellular amastigotes demonstrated that all genes shown in Table 3 (i.e., having their expression down-regulated in knockout cell lines compared to wild type cells) are up-regulated in epimastigotes compared to the other forms (Supplementary fig S2). Thus, the results of RNA-seq data indicate that TcRBP99 is an RNA binding factor responsible for the up-regulation of epimastigote-specific genes. It is tempting to speculate that by knocking out the expression of the gene encoding this RBP, expression of genes that are required for epimastigote proliferation are negatively affected and, because of that, the knockout cells are induced to differentiate into non-dividing metacyclic trypomastigotes.

Corroborating this assumption, among the transcripts that have their expression down-regulated in TcRBP99 knockout parasites is the TcCLB.506551.10 gene, annotated as a member of the carboxylate-transporter family, also known as PAD (protein associated with differentiation) (Fig 5A). Members of this family in *T. brucei*, which contain a conserved nodulin-like domain, are organized in tandem at the end of a polycistronic unit [27]. It has been shown that the expression of

459 the PAD2 gene is highly induced during differentiation of stumpy forms to
460 procyclic forms [27]. RNA-seq data described by Bellew and cols. (2017), as well
461 as by Li and cols. (2016) and Berna and cols. (2017) allowed us to determine that
462 the expression of several members this gene family is up-regulated in *T. cruzi*
463 epimastigotes not only in the CL Brener strain [7] but also in Y and Dm28c strains
464 (Table S2) [4, 5]. Quantitative qPCR confirmed that the expression of the
465 TcCLB.506551.10 gene is reduced 40% and 75% in knockout cell lines (clone 1
466 and clone 2, respectively) and that mRNA expression for this gene is restored in
467 both addback cell lines (Fig 5A).

468 Members from a second gene family whose expression was negatively affected
469 by the knockout of TcRBP99 encode amino acid transporters (Fig 5B). Again,
470 RNA-seq data from different *T. cruzi* strains showed that this family is up-
471 regulated epimastigotes (Table S2). Quantitative qPCR also confirmed that the
472 expression of the TcCLB.506153.10 gene is reduced about 5-fold in both cloned
473 knockout cells lines and is restored in the addback cell lines (Fig 5B). Finally,
474 members from the retrotransposon hot spot (RHS) proteins, were also identified
475 as a third family whose expression was affected by the knockout of TcRBP99.
476 This gene family is present as clusters of genes in subtelomeric regions of the *T.*
477 *cruzi* and *T. brucei* genomes, which encode 757 and 118 members, respectively
478 [2, 28]. Although they are highly variable, they possess a conserved central
479 region containing an ATP/GTP-binding motif and the RIME/ingi transposon
480 insertion site {Bringaud et al., 2012). Because of the large number of members in
481 this family and the fact that sequences encoding RHS are highly variable, it was
482 not possible to design primers that could amplify transcripts from one specific
483 copy of the RHS gene family to evaluate by qPCR the changes in transcript

484 levels. Besides RHS, PAD and amino acid transporter genes, the RNA-seq
485 analyses comparing WT and TcRBP99 knockout cell lines also identified 6
486 distinct sequences annotated as hypothetical genes in the *T. cruzi* genome
487 database.

488

489 **TcRBP99 binds to epimastigote-specific transcripts**

490 To identify mRNAs that may be targets of TcRBP99, we used a transfected cell
491 line expressing TcRBP99 with the HA tag to perform RNA immunoprecipitation
492 assays. Following immunoprecipitation with anti-HA antibodies, total RNA was
493 purified from immunocomplexes as well as from the supernatant containing RNAs
494 that were not bound to the HA tagged TcRBP99 and analysed by reverse
495 transcribed PCR. As shown in Fig 6A, western blot assays performed with total
496 protein extracts from wild type epimastigotes and transfected epimastigotes
497 showed a 20KDa corresponding to the TcRBP99::HA tagged protein only in
498 extracts from transfected parasites. The same band can also be detected in the
499 immunoprecipitated complex (eluate) but not in the flow through fraction. Note
500 that 25 and 50 KDa bands corresponding to IgG heavy and light chains are also
501 detected, their signals being stronger in the immunoprecipitated fraction. We
502 confirmed the interaction between TcRBP99 and the mRNA encoding PAD
503 (TcCLB.506551.10) by showing that a 1.7 kb band corresponding to the cDNA
504 derived from this mRNA can be amplified from RNA extracted from the
505 immunoprecipitated fractions obtained from TcRBP99::HA expressing parasites.
506 This PCR product, which was amplified from total RNA samples obtained from
507 WT as well from transfected parasites, was not present in samples derived from
508 immunoprecipitated complexes obtained from WT cells (Fig 6B). cDNAs

509 synthesized in the absence of the reverse transcriptase enzyme, used as a
510 negative control of the PCR reaction did not show PCR products either in the total
511 RNA or in the immunoprecipitated fraction derived from transfected parasites (Fig
512 6C). To test the specificity of the interaction between TcRBP99 and PAD
513 transcripts, we perform a similar immunoprecipitation assay with a *T. cruzi*
514 transfected cell line expressing the AGO/PIWI-like protein tagged with HA tag
515 {Garcia-Silva et al., 2010; Viviane G. Silva, unpuplished}. As shown in Fig 6C-D,
516 no PCR product corresponding to the PAD transcript was amplified from the
517 immunoprecipitated fraction obtained from cells expressing another HA-tagged
518 protein.

519 Taken together, the data presented here indicate that TcRBP99 is a trans-acting
520 factor that is abundantly expressed in epimastigotes and controls expression of
521 several genes that are also up-regulated in the insect stage of the *T. cruzi* life
522 cycle. By direct binding to its targets, TcRBP99 likely stabilizes mRNAs that
523 encode proteins involved with the epimastigote metabolism and, consequently,
524 allows the parasite to proliferate in the insect midgut. Our findings are in
525 agreement with recent data published by various groups showing drastic changes
526 in the *T. cruzi* transcriptome, that particularly affect the expression of genes
527 involved with the adaptation of the distinct metabolic profiles required for parasite
528 survival in the insect vector and the mammalian host. Additional efforts are
529 required to explore further the interactions between this RBP and its targets
530 mRNAs, the associated factors that participate in these interactions as well as
531 the mechanisms by which these interactions (or the lack of it) affect gene
532 expression. Together with the observation that TcRBP99 orthologues are present
533 only in trypanosomatids and the scarcity of studies in this parasite investigating

molecular mechanisms responsible for post-transcriptional control of gene expression, this report can be considered a significant step towards the understanding of an important aspect of host-parasite interaction that may help finding new methods to control human Chagas disease.

Acknowledgments

The authors wish to thank Renata Barbosa Peixoto for helpful help in cell culture and Gabriela Assis Burle Caldas for helping with microscopy analyses.

References

1. Brener Z: **Biology of Trypanosoma cruzi**. *Annual review of microbiology* 1973, **27**:347-382.

2. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G *et al*: **The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease**. *Science* 2005, **309**(5733):409-415.

3. Araujo PR, Teixeira SM: **Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in Trypanosoma cruzi: a review**. *Memorias do Instituto Oswaldo Cruz* 2011, **106**(3):257-266.

4. Li Y, Shah-Simpson S, Okrah K, Belew AT, Choi J, Caradonna KL, Padmanabhan P, Ndegwa DM, Temanni MR, Corrada Bravo H *et al*:

Transcriptome Remodeling in Trypanosoma cruzi and Human Cells during Intracellular Infection. *PLoS pathogens* 2016, **12**(4):e1005511.

5. Berna L, Chiribao ML, Greif G, Rodriguez M, Alvarez-Valin F, Robello C: **Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in Trypanosoma cruzi**. *PeerJ* 2017, **5**:e3017.

6. Houston-Ludlam GA, Belew AT, El-Sayed NM: **Comparative Transcriptome Profiling of Human Foreskin Fibroblasts Infected with the Sylvio and Y Strains of Trypanosoma cruzi**. *PloS one* 2016, **11**(8):e0159197.

7. Belew AT, Junqueira C, Rodrigues-Luiz GF, Valente BM, Oliveira AER, Polidoro RB, Zuccherato LW, Bartholomeu DC, Schenkman S, Gazzinelli RT *et al*: **Comparative transcriptome profiling of virulent and non-virulent Trypanosoma cruzi underlines the role of surface proteins during infection**. *PLoS pathogens* 2017, **13**(12):e1006767.

8. De Gaudenzi J, Frasch AC, Clayton C: **RNA-binding domain proteins in Kinetoplastids: a comparative analysis**. *Eukaryotic cell* 2005, **4**(12):2106-2114.

9. Guerra-Slompo EP, Probst CM, Pavoni DP, Goldenberg S, Krieger MA, Dallagiovanna B: **Molecular characterization of the Trypanosoma cruzi specific RNA binding protein TcRBP40 and its associated mRNAs**. *Biochemical and biophysical research communications* 2012, **420**(2):302-307.

10. Kramer S, Kimblin NC, Carrington M: **Genome-wide in silico screen for CCCH-type zinc finger proteins of Trypanosoma brucei, Trypanosoma cruzi and Leishmania major**. *BMC genomics* 2010, **11**:283.

11. Caro F, Bercovich N, Atorrasagasti C, Levin MJ, Vazquez MP: **Trypanosoma cruzi: analysis of the complete PUF RNA-binding protein family**. *Experimental parasitology* 2006, **113**(2):112-124.

12. Clayton CE: **Networks of gene expression regulation in Trypanosoma brucei**. *Molecular and biochemical parasitology* 2014, **195**(2):96-106.

13. De Gaudenzi JG, Carmona SJ, Aguero F, Frasch AC: **Genome-wide analysis of 3'-untranslated regions supports the existence of post-transcriptional regulons controlling gene expression in trypanosomes**. *PeerJ* 2013, **1**:e118.

14. Morking PA, Rampazzo Rde C, Walrad P, Probst CM, Soares MJ, Gradia DF, Pavoni DP, Krieger MA, Matthews K, Goldenberg S *et al*: **The zinc finger protein TcZFP2 binds target mRNAs enriched during Trypanosoma cruzi metacyclogenesis**. *Memorias do Instituto Oswaldo Cruz* 2012, **107**(6):790-799.

15. Dallagiovanna B, Perez L, Sotelo-Silveira J, Smircich P, Duhagon MA, Garat B: **Trypanosoma cruzi: molecular characterization of TcPUF6, a Pumilio protein**. *Experimental parasitology* 2005, **109**(4):260-264.

16. Perez-Diaz L, Correa A, Moretao MP, Goldenberg S, Dallagiovanna B, Garat B: **The overexpression of the trypanosomatid-exclusive TcRBP19 RNA-binding protein affects cellular infection by**

Trypanosoma cruzi. *Memorias do Instituto Oswaldo Cruz* 2012, **107**(8):1076-1079.

17. Perez-Diaz L, Silva TC, Teixeira SM: **Involvement of an RNA binding protein containing Alba domain in the stage-specific regulation of beta-amastin expression in Trypanosoma cruzi**. *Molecular and biochemical parasitology* 2017, **211**:1-8.

18. Alcantara MV, Kessler RL, Goncalves REG, Marliere NP, Guarneri AA, Picchi GFA, Fragoso SP: **Knockout of the CCCH zinc finger protein TcZC3H31 blocks Trypanosoma cruzi differentiation into the infective metacyclic form**. *Molecular and biochemical parasitology* 2018, **221**:1-9.

19. Camargo EP: **Growth and Differentiation in Trypanosoma Cruzi. I. Origin of Metacyclic Trypanosomes in Liquid Media**. *Revista do Instituto de Medicina Tropical de Sao Paulo* 1964, **6**:93-100.

20. Shaw AK, Kalem MC, Zimmer SL: **Mitochondrial Gene Expression Is Responsive to Starvation Stress and Developmental Transition in Trypanosoma cruzi**. *mSphere* 2016, **1**(2).

21. DaRocha WD, Silva RA, Bartholomeu DC, Pires SF, Freitas JM, Macedo AM, Vazquez MP, Levin MJ, Teixeira SM: **Expression of exogenous genes in Trypanosoma cruzi: improving vectors and electroporation protocols**. *Parasitology research* 2004, **92**(2):113-120.

22. Schumann Burkard G, Jutzi P, Roditi I: **Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters**. *Molecular and biochemical parasitology* 2011, **175**(1):91-94.

23. Grazielle-Silva V, Zeb TF, Bolderson J, Campos PC, Miranda JB, Alves CL, Machado CR, McCulloch R, Teixeira SM: **Distinct Phenotypes Caused by Mutation of MSH2 in Trypanosome Insect and Mammalian Life Cycle Forms Are Associated with Parasite Adaptation to Oxidative Stress**. *PLoS neglected tropical diseases* 2015, **9**(6):e0003870.

24. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method**. *Methods* 2001, **25**(4):402-408.

25. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.

26. Anders S, Pyl PT, Huber W: **HTSeq--a Python framework to work with high-throughput sequencing data**. *Bioinformatics* 2015, **31**(2):166-169.

27. Dean S, Marchetti R, Kirk K, Matthews KR: **A surface transporter family conveys the trypanosome differentiation signal**. *Nature* 2009, **459**(7244):213-217.

28. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B *et al*: **The genome of the African trypanosome Trypanosoma brucei**. *Science* 2005, **309**(5733):416-422.

Figure legends

Figure 1. RNA binding proteins up-regulated in epimastigotes.

RNA-Seq data generated using mRNA extracted from epimastigotes, amastigotes 60 hours post-infection and trypomastigotes were used to identify the RBP up-regulated in epimastigotes. All 253 T. cruzi RBP, previously identify by Belew et.al, were used. **(A)** Scatterplot comparing epimastigotes RBP expression with amastigote (upper panel) and trypomastigote (bottom panel). Circles indicate the transcripts up-regulated in both contrasts and the IDs are identify by the numbers. **(B)** Normalized reads counts corresponding the RBPs up-regulated in epimastigotes compared with mammalian stages.

Figure 2. *In silico* characterization of gene encoding for the TcRBP99.

Sequence of TcRBP99 orthologs genes were extracted from TritrypDB data bank and used to perform **(A)** multiple alignment and **(B)** phylogenetic tree construction. **(C)** Epimastigotes forms over-expressing the TcRBP99 with HA tag were used to localize the protein using antibodies anti HA following by antibodies anti-rat conjugated with Alexa 488. Nuclei and kinetoplast were stained with DAPI. The syntenic genes are represented by the tritrypDB identifiers: TcCLB.506739.99 (*T. cruzi* Esmeraldo), TcCLB.510819.119 (*T. cruzi* Non Esmeraldo), TvY486_0501060 (*T. vivax*), TcIL3000_5_1490 (*T. congolense*), TevSTIB805.5.1760 (*T. evansi*), Tb927.5.1570 (*T. brucei* 927), Tb427.05.1570 (*T. brucei* 427), Tbg972.5.2150 (*T. brucei gambiense*), LbrM.15.0140 (*L. braziliensis*), LtaP15.0130 (*L. tarentolae*), LmxM.15.0140 (*L. mexicana*), LinJ.15.0140 (*L. infantum*), LdBPK_150140.1 (*L. danovani*), LMJLV39_150006500 (*L. major*), LmjF.15.0140 (*L. major Friedlin*), CFAC1_060021400 (*C. fasciculata*), LpyrH10_38_0260 (*L. pyrrhocoris*),

675 Lsey_0162_0120 (*L. seymouri*), DQ04_00711020 (*T. grayi*), TRSC58_03397 (*T.*
676 *rangeli*)

677 **Figure 3. Characterization of over expression of TcRBP99 in epimastigotes**
678 **forms.**

679 **(A)** TcRBP99 coding region with HA tag in C-terminus were amplified from CL
680 Brener genome and cloned in pROCK_Neo T. cruzi vector expression. This
681 plasmid was linearized and then used in epimastigotes transfection that were
682 selected with neomycin drug. **(B)** Western blot with total protein extract from wild
683 type, population and clone cell lines over expressing the TcRBP99. **(C)** qPCR
684 using specific primers for TcRBP99. **(D)** Growth curve of WT, clone over
685 expressing the TcRBP99 and empty pROCK vector. **(E)** Metacyclogenesis were
686 induced transferring parasites to RPMI medium. Percent of metacyclic
687 trypomastigotes were evaluated in cell lines corresponding to WT, clone over
688 expressing the TcRBP99 and empty pROCK vector.

689 **Figure 4. Characterization of TcRBP99 knockout cell lines.**

690 **(A)** Schematic representation of knockout cell lines generation by homologous
691 recombination. The first allele was interrupt by sequencing containing the gene
692 corresponding to neomycin gene flanked by TcRBP99 sequences. Cell line
693 resistant to neomycin was used in transfection with the same sequence, but now
694 with hygromycin gene. Knockout cell lines were select with neomycin and
695 hygromycin drugs. pROCK vector expression contain puromycin resistance were
696 used to obtain add back cell lines, transfecting KO parasites with these *T. cruzi*
697 vector expression. **(B)** PCR performed with different primers combination to verify
698 the correct sequences integration. P1 and P2 amplify neomycin integration in
699 TcRBP99 locus (1,648 bp); P3 and P4 amplify hygromycin integration in

TcRBP99 locus (2,479 bp); P5 and P6 amplify TcRBP99 coding region (522 bp, 2,363 bp, interrupt by neomycin sequence, and 2,515 bp, interruption by hygromycin sequence); P7 and P8 amplify the sequence corresponding to puromycin (640 bp). **(C)** qPCR using specific primers for TcRBP99. **(D)** Growth curve of WT, knockout TcRBP99 clones, add back clones, and empty pROCK vector. **(E)** Metacyclogenesis were induced transferring parasites to RPMI medium. Percent of metacyclic trypomastigotes were evaluated in cell lines corresponding to WT, knockout TcRBP99 clones, add back clones, and empty pROCK vector.

Figure 5. Expression of genes differently expressed between TcRBP99 KO cell lines and wild type parasite.

qPCRs were performed to determine the mRNA expression of **(A)** protein associated with differentiation and **(B)** amino acid transporter in TcRBP99 hemi-knockout, TcRBP99 KO and add back cell lines.

Figure 6. Interaction between mRNA encoding for protein associated with differentiation and TcRBP99.

Parasites over expressing the TcRBP99 with HA tag were in immunoprecipitation assay with beads containing antibodies anti-HA. **(A)** Western blot with total protein extract of input, flow through and eluate fractions, incubated with anti HA antibodies. mRNA were extracted from all fractions, cDNA were synthetase in the **(B)** presence or **(C)** absent of reverse transcriptase and then coding region corresponding to the protein associated with differentiation were amplify from all fraction (1,845 bp). Another *T. cruzi* protein with HA tag were used as negative control. As the same after mRNA extraction cDNA were synthetase in the **(D)**

724 presence or **(E)** absent of reverse transcriptase and coding region corresponding
725 of PAD gene were amplify from all fraction.

726

727 Supporting Information

728 **Figure S1. Global statistical assessment of biological replicates**

729 Principal Component Analysis (PCA) plots corresponding to 3 replicates of wild
730 type parasites and 4 replicates, 2 of each clone, of KO cell lines.

731 **Figure S2. Global expression of families differently expressed in KO cell**
732 **lines compared with WT parasites.**

733 RNA-Seq data from epimastigotes, amastigotes 60 hours post infection,
734 amastigotes 96 hours post infection and trypomastigotes were used to determine
735 the global expression. Heatmap representing the global expression of the
736 members which constitute the amino acid transporters, the protein associated
737 with differentiation (PAD) and the retrotransposons hot spot (RHS) families.

738 **Figure S3. Negative control of immunoprecipitation.**

739 Parasites over expressing other T. cruzi protein containing HA tag (Tc::HA) were
740 used in immunoprecipitation assay with beads containing antibodies anti-HA. **(A)**
741 Western blot with protein extract of input, flow through and eluate fractions,
742 incubated with anti HA antibodies. **(B)** mRNA were extracted from all fractions
743 and then cDNA were synthetase in the presence or absent of reverse
744 transcriptase. Coding region corresponding to the protein associated with
745 differentiation were amplify from all fraction (1,845 bp).

746 **Table S1. Primer sequences**

Table S2. Expression of DE genes in others *T. cruzi* strains.

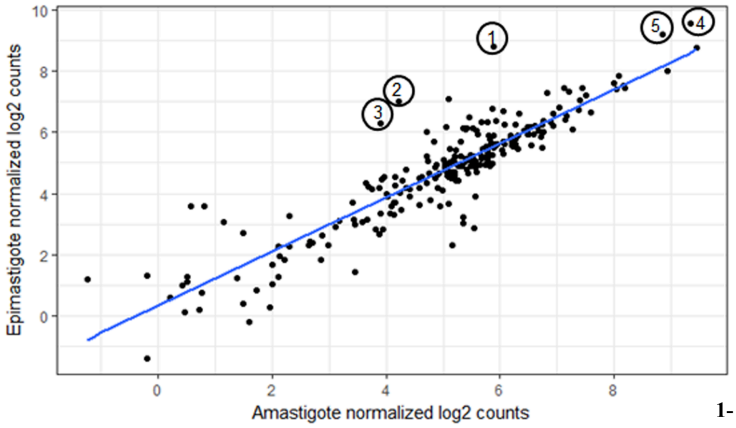
Expression of genes encoding for protein associated with differentiation and amino acid transporter, differently expressed in KO parasites compared with WT, in transcriptome of Y and Dm28c *T. cruzi* strains previously published.

Table 1. Gene ontology (GO) enriched in CL Brener epimastigotes compared to amastigotes and trypomastigotes.

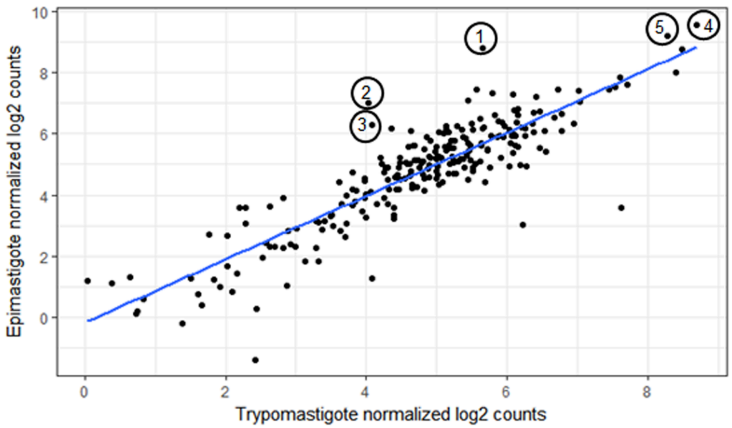
GO Identifier	GO Term	P-value	Genes DE	Total genes
Biological process				
GO:0055114	oxidation-reduction process	3,22E-24	121	386
GO:0008152	metabolic process	2,8E-09	73	302
GO:0055085	transmembrane transport	2,44E-07	42	151
GO:0006414	translational elongation	1,6E-07	35	117
GO:0006457	protein folding	0,05442	20	115
GO:0005975	carbohydrate metabolic process	4,19E-05	19	59
GO:0006629	lipid metabolic process	0,000176	18	60
GO:0009058	biosynthetic process	0,000198	16	51
GO:0015991	ATP hydrolysis coupled proton transport	1,6E-05	14	34
GO:0006520	cellular amino acid metabolic process	1,33E-08	13	19
Molecular function				
GO:0016491	oxidoreductase activity	9,91E-18	75	22
GO:0003824	catalytic activity	0,000471	70	388
GO:0016787	hydrolase activity	0,000689	34	160
GO:0003746	translation elongation factor activity	4,25E-06	30	106
GO:0030170	pyridoxal phosphate binding	3,46E-08	22	52
GO:0009055	electron transfer activity	5,24E-10	21	40
GO:0016740	transferase activity	4,2E-06	19	52
GO:0020037	heme binding	2,85E-05	16	44
GO:0004375	glycine dehydrogenase (decarboxylating) activity	2,1E-06	15	32
GO:0004198	calcium-dependent cysteine-type endopeptidase activity	0,000478	11	51

Figure 1

A



1- TcCLB.506739.99
2- TcCLB.506931.4
3- TcCLB.510131.44
4- TcCLB.511727.290
5- TcCLB.511727.270



B

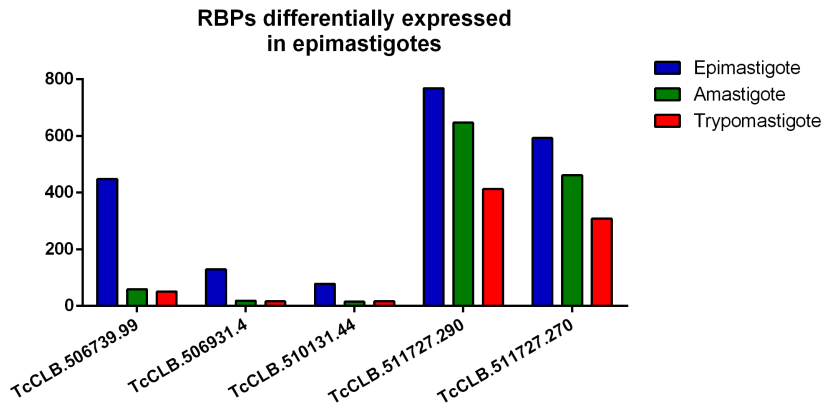
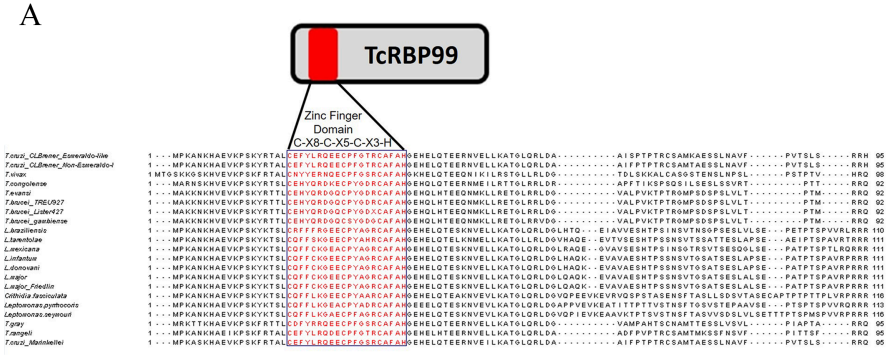
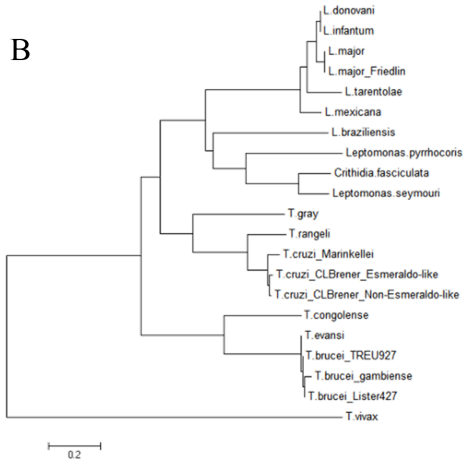


Figure 2

A


TcRBP99
Zinc Finger
Domain
C-X8-C-X5-C-X3-H

B


L.donovani
L.infantum
L.major
L.major_Friedlin
L.tarentolae
L.mexicana
L.braziliensis
Leptomonas.pyrrhocoris
Crithidia.fasciculata
Leptomonas.seymouri
T.gray
T.rangeli
T.cruzi_Marinkellei
T.cruzi_CLBrener_Esmeraldo-like
T.cruzi_CLBrener_Non-Esmeraldo-like
T.congolense
T.evansi
T.brucei_TREU927
T.brucei_gambiense
T.brucei_Lister427
T.vivax
0.2

C

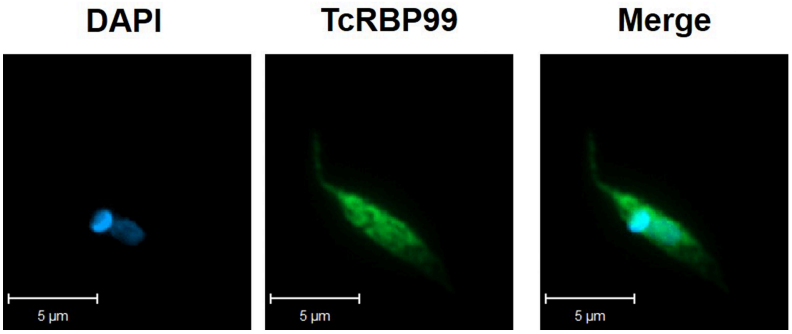

DAPI
TcRBP99
Merge
5 µm
5 µm
5 µm

Figure 3

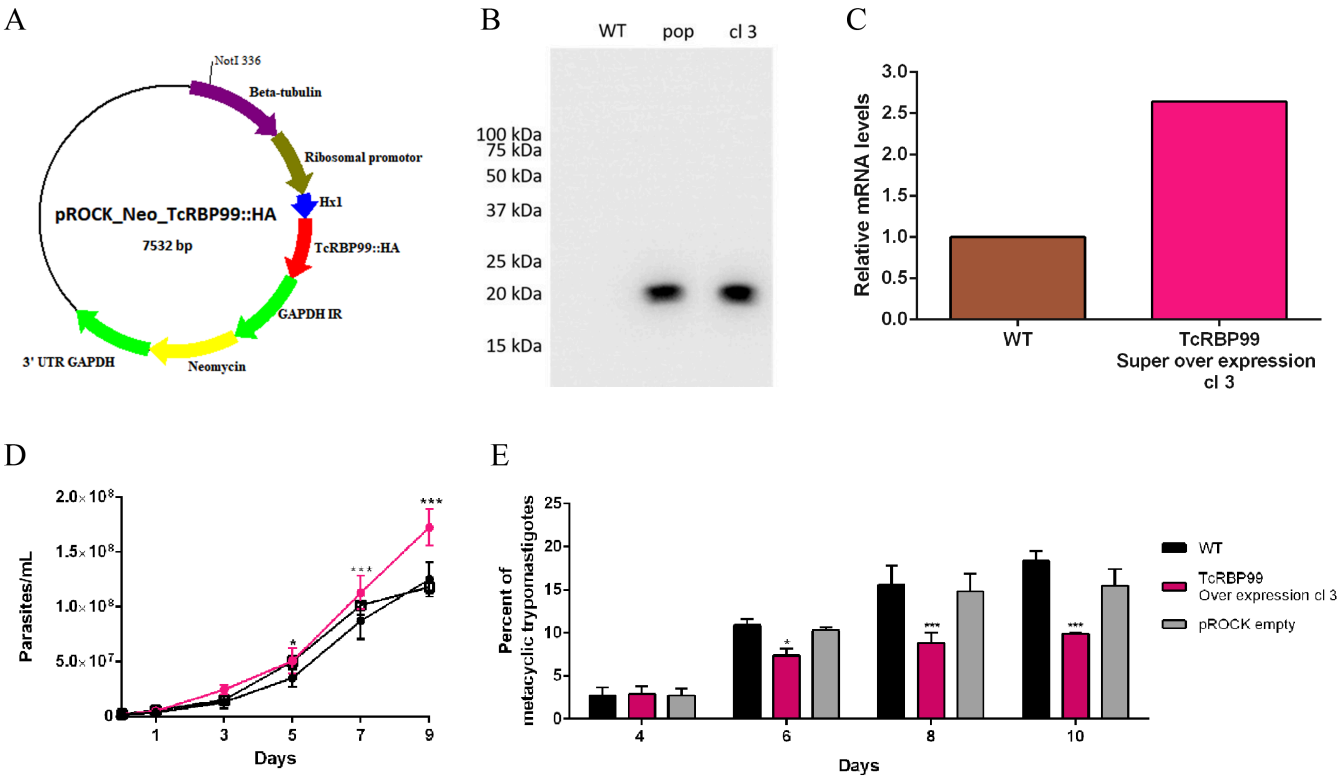
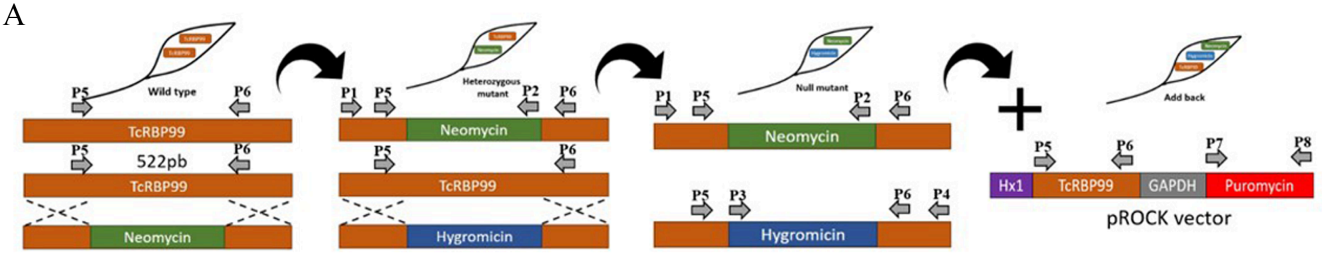
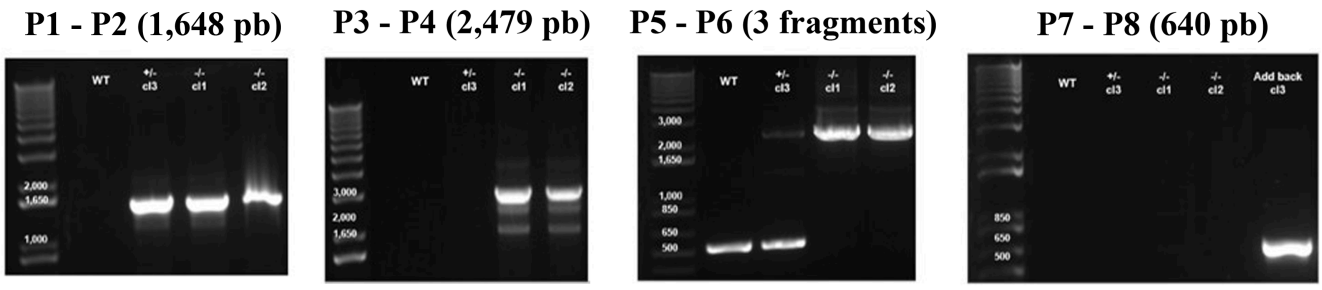

A
NotI 336
Beta-tubulin
Ribosomal promotor
Hx1
pROCK_Neo_TcRBP99::HA
7532 bp
TcRBP99::HA
GAPDH IR
3' UTR GAPDH
Neomycin
B
WT
pop
cl 3
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
20 kDa
15 kDa
C
Relative mRNA levels
3.0
2.5
2.0
1.5
1.0
0.5
0.0
WT
TcRBP99
Super over expression
cl 3
D
Parasites/mL
2.0× 10^8
1.5× 10^8
1.0× 10^8
5.0× 10^7
0
1
3
5
7
9
Days
E
Percent of metacyclic trypomastigotes
25
20
15
10
5
0
4
6
8
10
Days
WT
TcRBP99
Over expression cl 3
pROCK empty

Figure 4

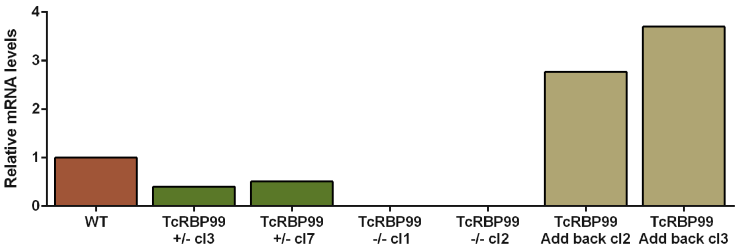
A



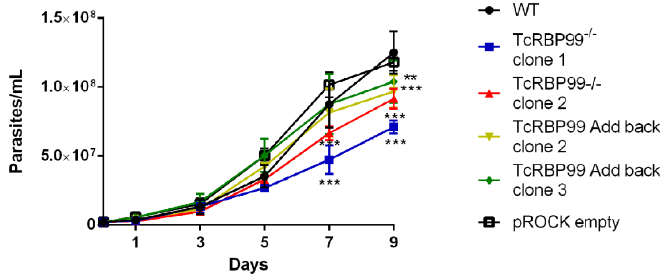
B



C



D



E

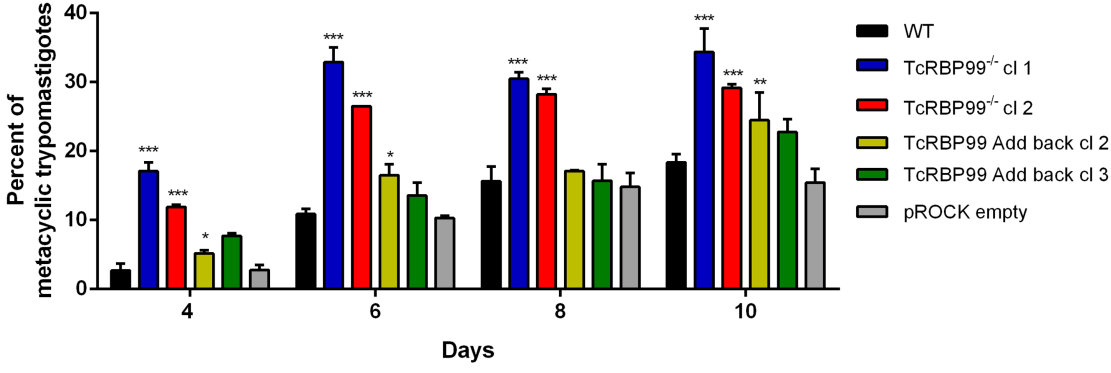


Table 2 – Summary of total reads obtain in the RNA-seq and mapped in *T. cruzi* CL Brener genome

Cell line	Total reads	Reads after trimming	Reads mapped (%)
Wild type	3,294,361	2,709,211	72.5%
Wild type	3,609,438	2,862,761	74.9%
Wild type	3,469,573	2,626,697	75.0%
Knockout	3,513,887	2,741,084	70.2%
Knockout	3,895,258	3,088,360	72.9%
Knockout	4,143,756	3,274,477	73.1%
Knockout	3,521,679	2,820,681	74.4%

Table 3 – Genes differently expressed (DE) comparing knockout cell lines and wild type parasites

ID	Annotation in CL Brener genome	log2 fold change	p-value
TcCLB.509559.20	retrotransposon hot spot protein (RHS, pseudogene)	-2,86223	1,06E-35
TcCLB.505943.110	retrotransposon hot spot protein (RHS, pseudogene)	-2,74450	1,52E-32
TcCLB.503733.80	hypothetical protein, conserved	-1,50603	1,42E-19
TcCLB.506153.10	amino acid transporter	-1,26987	3,04E-08
TcCLB.506551.10	protein associated with differentiation 8	-1,14790	1,01E-14
TcCLB.504173.60	hypothetical protein	-1,14374	1,04E-07
TcCLB.504163.40	hypothetical protein	-1,12380	3,65E-05
TcCLB.506825.40	hypothetical protein	-1,12250	6,19E-05
TcCLB.507659.10	hypothetical protein	-1,05995	4,13E-06
TcCLB.511185.150	hypothetical protein (pseudogene)	-1,01498	0,000622783
TcCLB.511843.80	retrotransposon hot spot (RHS) protein	-1,01447	2,80E-05

Genes DE: p-value <0.05 and log2FC <1

Figure 5

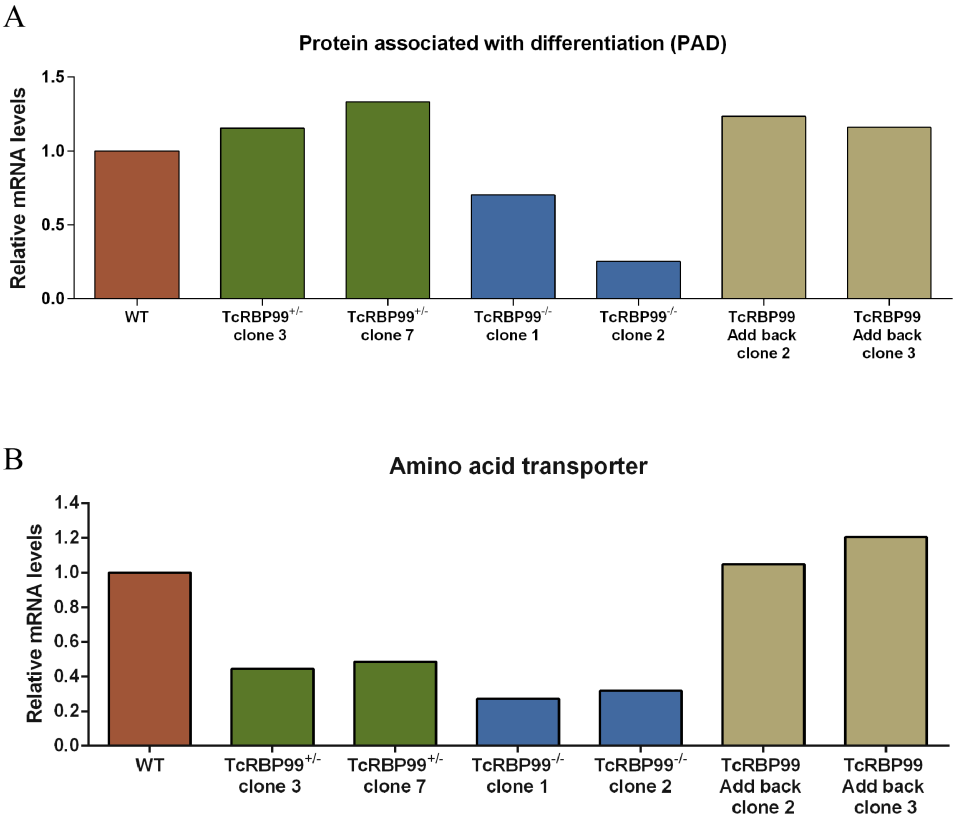
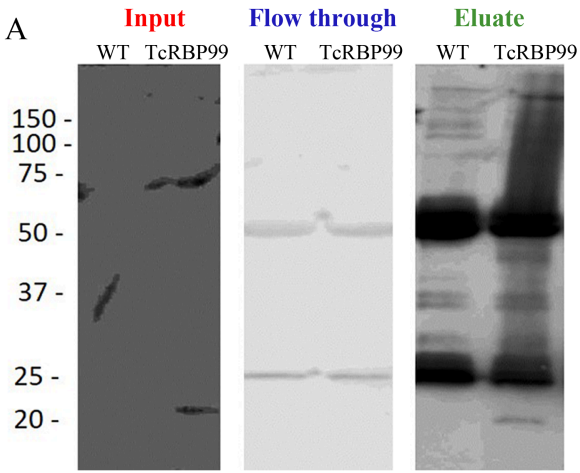
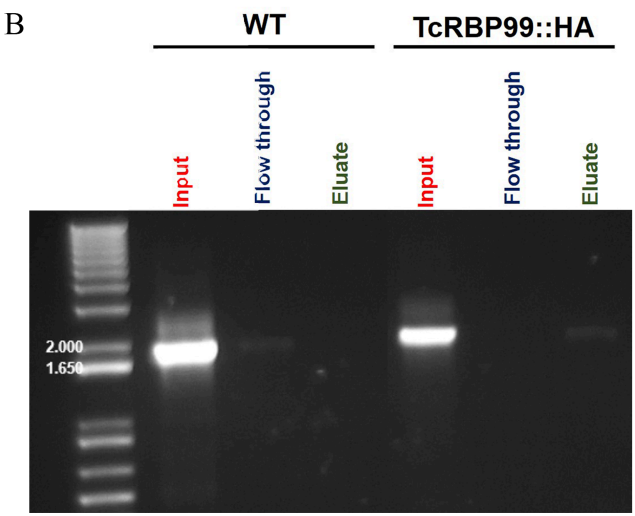
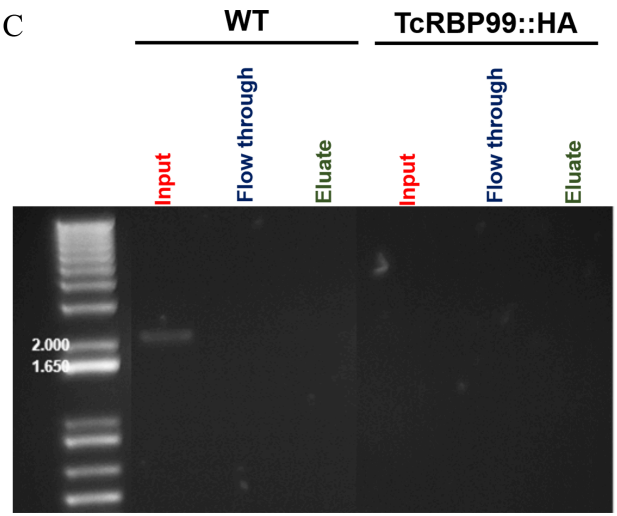
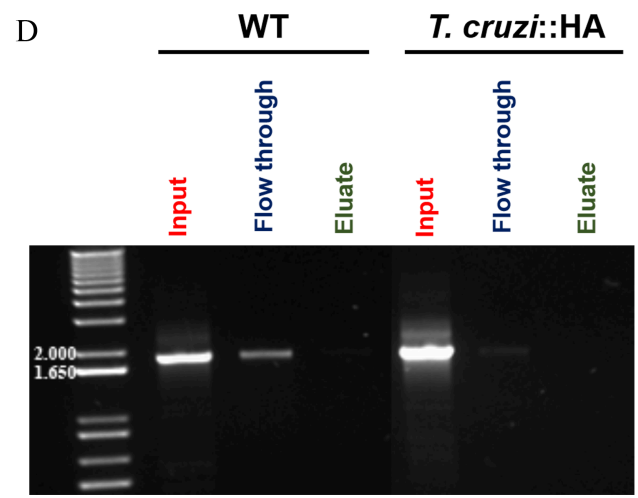

A
Protein associated with differentiation (PAD)
Relative mRNA levels
1.5
1.0
0.5
0.0
WT
TcRBP99+/- clone 3
TcRBP99+/- clone 7
TcRBP99-/- clone 1
TcRBP99-/- clone 2
TcRBP99 Add back clone 2
TcRBP99 Add back clone 3
B
Amino acid transporter
Relative mRNA levels
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
WT
TcRBP99+/- clone 3
TcRBP99+/- clone 7
TcRBP99-/- clone 1
TcRBP99-/- clone 2
TcRBP99 Add back clone 2
TcRBP99 Add back clone 3

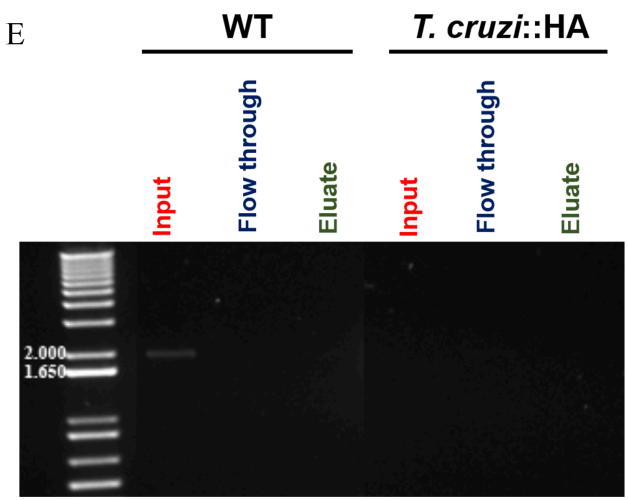
Figure 6


A
Input
Flow through
Eluate
WT
TcRBP99
WT
TcRBP99
WT
TcRBP99
150 -
100 -
75 -
50 -
37 -
25 -
20 -

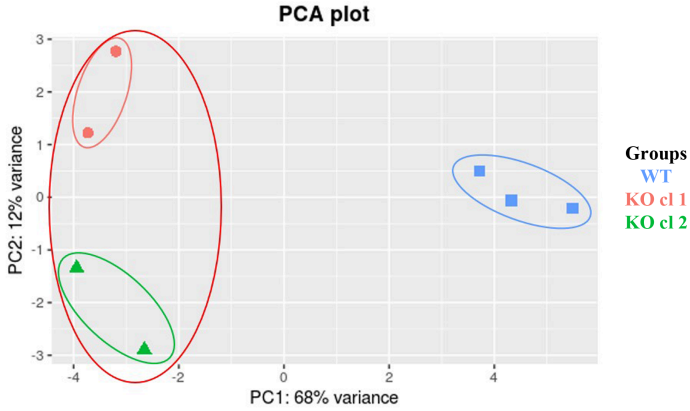

B
WT
TcRBP99::HA
Input
Flow through
Eluate
Input
Flow through
Eluate
2.000
1.650


C
WT
TcRBP99::HA
Input
Flow through
Eluate
Input
Flow through
Eluate
2.000
1.650

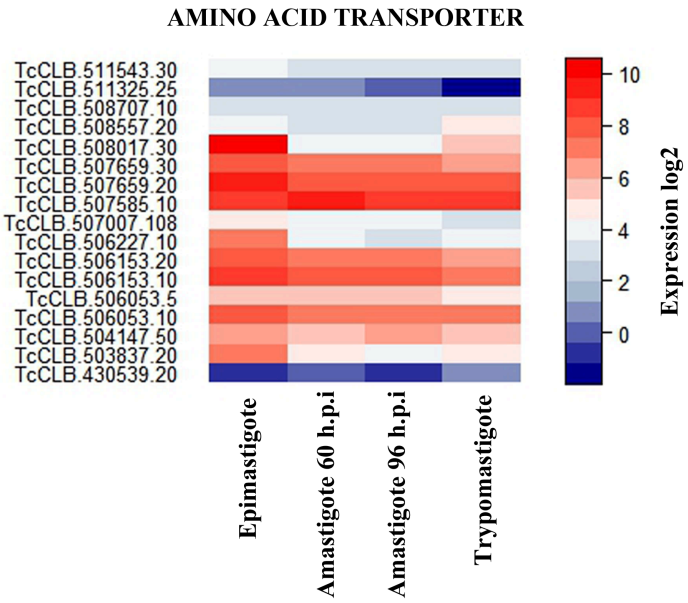

D
WT
T. cruzi::HA
Input
Flow through
Eluate
Input
Flow through
Eluate
2.000
1.650

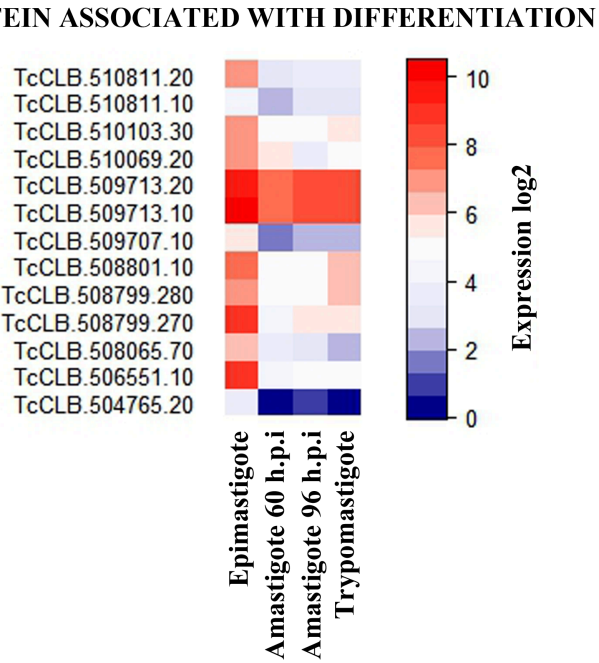

E
WT
T. cruzi::HA
Input
Flow through
Eluate
Input
Flow through
Eluate
2.000
1.650

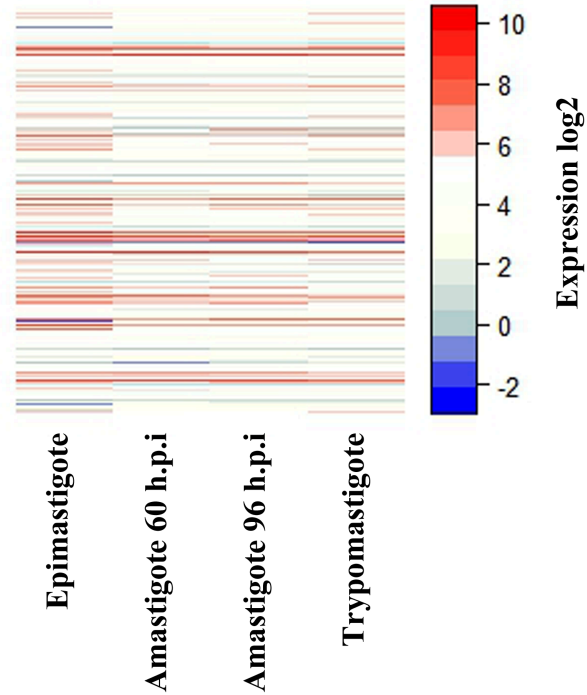
Suplementary Figure 1



Suplementary Figure 2







Apêndice V: Lista das sequencias contendo motivos de ligação a RNA identificadas no genoma do clone CL Brener do *T. cruzi*

Gene ID	TriTrypDB anotação	Domínio	Descrito por
TcCLB.397937.10 TcCLB.508625.160	Pumilio/PUF RNA binding protein 1	PUF	4
TcCLB.503989.10 TcCLB.401469.10	Zinc finger protein 2	Zinc Finger	3
TcCLB.411427.10	Hypothetical protein	Zinc Finger	3
TcCLB.507711.40 TcCLB.424123.40	RNA-binding protein	RRM	1 ,2 1
TcCLB.508787.30 TcCLB.468005.9	Pumilio/PUF RNA binding protein 3	PUF	4
TcCLB.469785.40 TcCLB.507025.50	RNA-binding protein	RRM	1 1 ,2
TcCLB.507025.60	RNA-binding protein	RRM	1,2
TcCLB.503419.50 TcCLB.509105.90	MRB1-associated protein	RRM	1,2 1
TcCLB.506859.204 TcCLB.503567.9	Hypothetical protein	Zinc Finger	3
TcCLB.503577.20 TcCLB.510943.60	U2 splicing auxiliary factor	Zinc Finger	1,3
TcCLB.503619.20 TcCLB.511647.40	Hypothetical protein	RRM	1 1,2
TcCLB.503683.11 TcCLB.509999.140	Double RNA binding domain protein 6A	RRM	1,2 1
TcCLB.503683.30 TcCLB.509999.120	RNA-binding protein	RRM	1 1,2
TcCLB.509581.50 TcCLB.503709.10	Hypothetical protein	RRM	1 1,2
TcCLB.508569.90 TcCLB.503733.50	Hypothetical protein	RRM	1,2 1
TcCLB.503757.30 TcCLB.503719.39	Hypothetical protein	PUF	
TcCLB.503795.10 TcCLB.506679.10	Hypothetical protein	Zinc Finger	3
TcCLB.506563.10 TcCLB.503869.40	Pumilio/PUF RNA binding protein 9	PUF	4
TcCLB.503897.90 TcCLB.509561.110	Hypothetical protein	RRM	1 1,2
TcCLB.503897.150 TcCLB.509561.39	Hypothetical protein	Zinc Finger	3
TcCLB.507093.63 TcCLB.503917.7	RNA-binding protein	RRM	2
TcCLB.507037.20 TcCLB.508707.80	RNA-binding protein	RRM	1,2 1
TcCLB.508209.39	Hypothetical protein	RRM	1,2

TcCLB.503919.30			1
TcCLB.509591.50	RNA-binding protein 34	RRM	1,2
TcCLB.503999.90			1
TcCLB.504089.60	Hypothetical protein	Alba	
TcCLB.504001.10			
TcCLB.504089.70	Hypothetical protein	Alba	
TcCLB.504001.20			
TcCLB.504005.6	RNA-binding protein	RRM	1
TcCLB.511481.55			1,2
TcCLB.504071.80	Nuclear cap binding protein	RRM	1,2
TcCLB.507089.70	RNA-binding protein	RRM	1
TcCLB.504085.30			1,2
TcCLB.507089.30	Zinc finger protein	Zinc Finger	3
TcCLB.504085.70			
TcCLB.511109.130	Heterogeneous nuclear ribonucleoprotein H/F	RRM	1,2
TcCLB.504157.10			1
TcCLB.504165.20	Hypothetical protein	RRM	1,2
TcCLB.508981.20			1
TcCLB.508145.30	RNA-binding protein	RRM	1
TcCLB.504243.10			1,2
TcCLB.504431.90	U2 small nuclear ribonucleoprotein B	RRM	1
TcCLB.507951.140			1,2
TcCLB.511577.30	Hypothetical protein	Zinc Finger	3
TcCLB.504797.110			
TcCLB.504929.5	Zinc finger CCCH domain containing protein 11	Zinc Finger	3
TcCLB.507305.40			
TcCLB.505007.10	RNA-binding protein	RRM	1,2
TcCLB.510007.30			1
TcCLB.505165.10	RNA-binding protein	RRM	1
TcCLB.510295.59	Hypothetical protein	Zinc Finger	3
TcCLB.510297.10	Hypothetical protein	Zinc Finger	
TcCLB.506009.10			3
TcCLB.506127.20	mRNA export factor MEX67	Zinc Finger	3
TcCLB.506211.60	RNA-binding protein	Zinc Finger	3
TcCLB.508895.60			
TcCLB.506211.70	RNA-binding protein	Zinc Finger	3
TcCLB.508895.50			
TcCLB.510101.200	Zinc finger ccch and cchc domain-containing protein	Zinc Finger	
TcCLB.506297.130			
TcCLB.506399.40	Hypothetical protein	RRM	1
TcCLB.509243.20			
TcCLB.506425.60	Hypothetical protein	RRM	1,2
TcCLB.506565.12	RNA-binding protein	RRM	2
TcCLB.511837.129			
TcCLB.506565.4	RNA-binding protein	RRM	2

TcCLB.506565.8	RNA-binding protein	RRM	
TcCLB.506625.70	RNA-binding protein	RRM	1,2
TcCLB.506625.80 TcCLB.508349.39	Double RNA binding domain protein 3	RRM	1,2 1
TcCLB.506825.10 TcCLB.506681.10	Double RNA binding domain protein 12	RRM	1 1,2
TcCLB.506693.30 TcCLB.508153.680	RNA-binding protein 6	RRM	1,2 1
TcCLB.506733.140 TcCLB.509229.90	Hypothetical protein	Zinc Finger	3
TcCLB.506739.99 TcCLB.510819.119	Zinc finger (CCCH type) protein	Zinc Finger	3
TcCLB.506773.130 TcCLB.508799.70	Pumilio-repeat, RNA-binding protein	PUF	
TcCLB.506795.10 TcCLB.509937.60	RNA-binding protein	RRM	1,2 1
TcCLB.506797.120 TcCLB.511907.100	Hypothetical protein	RRM	1,2 1
TcCLB.509167.30 TcCLB.506831.30	Hypothetical protein	RRM	1 1,2
TcCLB.510759.100 TcCLB.506999.120	Hypothetical protein	Zinc Finger	3
TcCLB.509551.60 TcCLB.507007.77	Hypothetical protein	Zinc Finger	3
TcCLB.508271.4	Hypothetical protein	Zinc Finger	3
TcCLB.508479.120	Pumilio/PUF RNA binding protein 8	PUF	
TcCLB.506859.80 TcCLB.511815.50	Hypothetical protein	Zinc Finger	3
TcCLB.506859.230 TcCLB.511817.10	Hypothetical protein	Zinc Finger	3
TcCLB.506859.240 TcCLB.511817.20	Hypothetical protein	Zinc Finger	3
TcCLB.510427.10 TcCLB.506883.120	Hypothetical protein	Zinc Finger	3
TcCLB.506885.200 TcCLB.510729.220	Hypothetical protein	Zinc Finger	3
TcCLB.506885.204 TcCLB.510729.210	Hypothetical protein	Zinc Finger	3
TcCLB.506885.70 TcCLB.508461.140	Polyadenylate-binding protein 1	RRM/PABP	1,2
TcCLB.506925.400	Hypothetical protein	RRM	1,2
TcCLB.506931.4 TcCLB.510131.44	Hypothetical protein	Zinc Finger	3
TcCLB.506933.50	Hypothetical protein	Zinc Finger	3

TcCLB.506945.210	Zinc finger CCCH domain-containing protein 47	Zinc Finger	3
TcCLB.503453.20	Hypothetical protein	Zinc Finger	3
TcCLB.506977.110	Hypothetical protein	Zinc Finger	3
TcCLB.506989.100	Triple RNA binding domain protein 3	RRM	1
TcCLB.510149.140			1,2
TcCLB.507049.199	Pumilio protein	PUF	
TcCLB.507093.220	RNA-binding protein	RRM	1,2
TcCLB.507093.229	U-rich RNA-binding protein UBP-2	RRM	1,2
TcCLB.507093.250	RNA-binding protein	RRM	1,2
TcCLB.508213.40	Hypothetical protein	RRM	1
TcCLB.507515.40			1,2
TcCLB.508213.20	Hypothetical protein	RRM	1
TcCLB.507515.60			1,2
TcCLB.508577.100	Pumilio/PUF RNA binding protein 5	PUF	4
TcCLB.507521.110			
TcCLB.510351.80	Hypothetical protein	Zinc Finger	3
TcCLB.507601.80			
TcCLB.507611.300	U1A small nuclear ribonucleoprotein	RRM	1,2
TcCLB.507723.120			1
TcCLB.507625.70	Hypothetical protein	Zinc Finger	3
TcCLB.507787.140			
TcCLB.511867.10	Hypothetical protein	Zinc Finger	3
TcCLB.507775.10			
TcCLB.507831.20	Hypothetical protein	Zinc Finger	3
TcCLB.511263.50			
TcCLB.507831.30	Hypothetical protein	Zinc Finger	3
TcCLB.511263.40			
TcCLB.507831.40	Hypothetical protein	Zinc Finger	3
TcCLB.511263.30			
TcCLB.507831.110	Pumilio/PUF RNA binding protein 2	PUF	4
TcCLB.511261.120			
TcCLB.507873.30	Double RNA binding domain protein 7	RRM	1,2
TcCLB.510689.60			1
TcCLB.507885.10	RNA-binding protein	RRM	1
TcCLB.507993.140	RNA-binding protein 29	RRM	1
TcCLB.511277.580			
TcCLB.508145.10	RNA-binding protein	RRM	1
TcCLB.508145.20	RNA-binding protein	RRM	1
TcCLB.509149.20	Present in the outer mitochondrial membrane proteome 35	Zinc Finger	3
TcCLB.508175.350			
TcCLB.508215.10	Hypothetical protein	Zinc Finger	3
TcCLB.508879.10			
TcCLB.508241.90	Hypothetical protein	Zinc Finger	3
TcCLB.511151.20			

TcCLB.508357.9 TcCLB.508355.330	Hypothetical protein	Zinc Finger	3
TcCLB.510143.80 TcCLB.508409.270	RNA-binding protein	RRM	1 1,2
TcCLB.510143.120 TcCLB.508409.310	Hypothetical protein	Zinc Finger	3
TcCLB.510755.120 TcCLB.508413.50	RNA-binding protein	RRM	1 1,2
TcCLB.511715.100 TcCLB.508445.99	Pumilio/PUF RNA binding protein 7	PUF	4
TcCLB.508461.320	U5 snRNA-associated splicing factor	RRM	
TcCLB.510877.150 TcCLB.508689.20	Eukaryotic translation initiation factor 3 subunit g	RRM	1 1,2
TcCLB.508693.40	Hypothetical protein	Zinc Finger	3
TcCLB.508831.66 TcCLB.511671.130	Hypothetical protein	Zinc Finger	3
TcCLB.509805.240 TcCLB.511267.20	Hypothetical protein	Zinc Finger	3
TcCLB.509805.230 TcCLB.511267.24	Hypothetical protein	Zinc Finger	3
TcCLB.508837.140 TcCLB.511383.30	u1 small nuclear ribonucleoprotein 70 kDa	RRM	1,2
TcCLB.511277.200	RNA-binding protein	RRM	1,2
TcCLB.511367.60	Lupus La protein homolog	RRM	1,2
TcCLB.511467.4	Hypothetical protein	Zinc Finger	3
TcCLB.511127.10 TcCLB.511481.70	RNA-binding protein 5	RRM	1 1,2
TcCLB.508879.80	Hypothetical protein	RRM	1,2
TcCLB.424195.9 TcCLB.508901.20	RNA-binding protein 4	RRM	1,2
TcCLB.508989.30 TcCLB.509569.120	RNA-binding protein	RRM	1 1,2
TcCLB.509965.180 TcCLB.509053.179	RNA-binding protein	RRM	1,2 1
TcCLB.509055.10	RNA-binding protein	RRM	1
TcCLB.509231.39 TcCLB.509719.69	Hypothetical protein	Zinc Finger	3
TcCLB.509233.210	Hypothetical protein	Zinc Finger	3
TcCLB.511621.50 TcCLB.509317.60	RNA-binding protein	RRM	1,2 1
TcCLB.463297.10 TcCLB.509395.40	Hypothetical protein	Zinc Finger	3
TcCLB.510073.30 TcCLB.509399.190	Pumilio/PUF RNA binding protein 4	PUF	4

TcCLB.510507.50 TcCLB.509461.100	RNA-binding protein	RRM	1 1,2
TcCLB.510105.33 TcCLB.509715.23	RNA-binding protein	RRM	1 1,2
TcCLB.509759.19	Pumilio/PUF RNA binding protein 1	PUF	4
TcCLB.511039.39 TcCLB.509791.90	Hypothetical protein	Zinc Finger	3
TcCLB.510125.10	Pumilio/PUF RNA binding protein 6	PUF	4
TcCLB.511555.40 TcCLB.510219.30	cleavage and polyadenylation specificity factor 30 kDa subunit	Zinc Finger	3
TcCLB.511511.3	Zinc finger protein ZFP1	Zinc Finger	3
TcCLB.511511.6	Zinc finger protein ZFP1	Zinc Finger	3
TcCLB.511741.40 TcCLB.511517.70	RNA-binding protein	RRM	1 1,2
TcCLB.511735.40 TcCLB.511521.20	Hypothetical protein	Zinc Finger	3
TcCLB.511727.160	Polypyrimidine tract-binding protein	RRM	1,2
TcCLB.511727.190	Hypothetical protein	RRM	
TcCLB.511727.270	RNA-binding protein	RRM	1,2
TcCLB.511727.290	RNA-binding protein	RRM	1,2
TcCLB.511735.88	Hypothetical protein	Zinc Finger	3
TcCLB.511807.130	Hypothetical protein	Zinc Finger	3
TcCLB.511807.160	Hypothetical protein	Zinc Finger	3
TcCLB.511871.110 TcCLB.511863.20	Hypothetical protein	RRM	1 1,2
TcCLB.511867.60	Hypothetical protein	Zinc Finger	3
TcCLB.511837.138	RNA-binding protein	RRM	
TcCLB.510311.50 TcCLB.510265.40	Hypothetical protein	RRM	1
TcCLB.510747.80 TcCLB.510657.160	Double RNA binding domain protein 9	RRM	1,2 1
TcCLB.510877.30	Hypothetical protein	Alba	
TcCLB.510877.40	Hypothetical protein	Alba	

1 - De Gaudenzi e colaboradores (2005)

2 - Guerra-Slompo e colaboradores (2012)

3 – Kramer e colaboradores (2010)

4 – Caro e colaboradores (2006)